

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/009927

International filing date: 24 March 2005 (24.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/556,393
Filing date: 24 March 2004 (24.03.2004)

Date of receipt at the International Bureau: 02 August 2005 (02.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1348097

UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

July 21, 2005

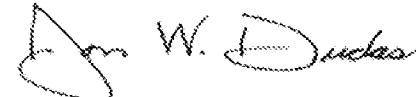
THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.

APPLICATION NUMBER: 60/556,393

FILING DATE: *March 24, 2004*

RELATED PCT APPLICATION NUMBER: PCT/US05/09927

Certified by



Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office



15992 U.S.PTO
032404

PTO/SB/16 (08-03)

Approved for use through 7/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EV331819218US

00/556393
7838 U.S. PTO
032404**INVENTOR(S)**

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Nelson R.	BARTON	San Diego, California
Eileen	O'DONOGHUE	San Diego, California
Gerhard	FREY	San Diego, California

Additional inventors are being named on the 1 separately numbered sheets attached hereto**TITLE OF THE INVENTION (500 characters max)**CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING
AND USING THEM

Direct all correspondence to:

CORRESPONDENCE ADDRESS

Customer Number: 25225

OR

Firm or
Individual Name

Address

City

Country

State

Zip

Telephone

Fax

ENCLOSED APPLICATION PARTS (check all that apply)

Specification Number of Pages

80

CD(s), Number

Drawing(s) Number of Sheets

Other

6 pages claims; 1 pg. abstract; Table 1
(specify): — 1 pg., 1 cover sheet

Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

Applicant claims small entity status. See 37 CFR 1.27.

A check or money order is enclosed to cover the filing fees.

FILING FEE
AMOUNT (\$)

The Director is hereby authorized to charge filing
fees or credit any overpayment to Deposit Account Number: 03-1952

80.00

Payment by credit card. Form PTO-2038 is attached.

The invention was made by an agency of the United States Government or under a contract with an agency of the
United States Government.

No Yes, the name of the U.S. Government agency
and the Government contract number are: _____

[Page 1 of 2]

Respectfully submitted,

SIGNATURE
TYPED OR
PRINTED NAME

Gregory P. Einhorn

TELEPHONE

(858) 720-5100

REGISTRATION NO.
(if appropriate)

38,440

Docket Number: 564463006801

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENTI hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EV331819218US,
in an envelope addressed to: MS Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-
1450, on the date shown below.Dated: 3/24/04

Signature:

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (08-03)

Approved for use through 07/31/06. OMB 0651-0032

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number 564463006801

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Ryan	SHORT	Rancho Santa Fe, California

[Page 2 of 2]

PROVISIONAL APPLICATION FOR UNITED STATES PATENT

CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM

Inventors: **Nelson R. Barton**, a citizen of the United States
residing at 10928 Sunset Ridge Dr., San Diego, CA 92131
Eileen O'Donoghue, a citizen of the United States
residing 1136 Missouri Street, San Diego, CA 92109
Gerhard Frey, a citizen of Germany,
residing at 13768 Via Cima Bella, San Diego, CA 92129
Ryan Short, a citizen of the United States,
residing in Rancho Santa Fe, CA,
and having a mailing address of : BOX 7214
Rancho Santa Fe, CA 92067-7214

Assignee: **Diversa Corporation**
4955 Directors Place
San Diego, California 92121 U.S.A.

Morrison & Foerster LLP
3811 Valley Centre Drive, Suite 500
San Diego CA 92130

ATTORNEY DOCKET: **DATE OF DEPOSIT:** March 24, 2004
56446-30068.01 / 115P02/ D1590 **EXPRESS MAIL NO.:** EV331819218US

CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM

TECHNICAL FIELD

5 This invention relates to nanotechnology, pharmacology and drug synthesis. In one aspect, the invention provides compositions and methods for the identification, separation or synthesis of proteins or ligands. In one aspect, the invention provides compositions comprising nanotubules of the invention and methods for making and using these nanotubules. In one aspect, the chimeric proteins and nanotubules of the 10 invention comprise a detectable moiety, e.g., a fluorescent protein. In one aspect, the invention provides compositions and methods for the selection and purification of chiral compositions from racemic mixtures. In one aspect, the invention provides chimeric cannulae polypeptides and methods for making and using them.

BACKGROUND

15 Enantiomers frequently display dramatically different pharmacological properties. As a result, use of single-enantiomer drugs may improve efficacy and reduce side effects. The United States Food and Drug Administration also recognizes the importance of understanding the pharmacological properties of each enantiomer. In order for a racemic drug to be registered, the biological activity of each purified enantiomer 20 must be characterized.

25 Cannulae A, or CanA, is a heat-resistant protein capable of forming nanotubules. CanA nanotubules are assembled from 21 kDa monomeric subunits that self-assemble in the presence of divalent cation into hollow rods with an outer diameter of approximately 25 nm and an inner diameter of approximately 20 nm, thus exhibiting molecular dimensions and an overall morphology not dissimilar to eukaryotic microtubules. CanA monomer expressed in *E. coli* is heat-stable. It can be rapidly 30 purified from bacterial extracts following heat treatment to remove the majority of the heat-labile host proteins. Following purification, the CanA monomer readily self-assembles into nanotubules in the presence of calcium and magnesium at elevated temperature. The assembled nanotubule structure contains 28 CanA monomers per turn arranged with a helical pitch. The CanA nanotubules are heat stable (up to 128°C) and

remain assembled in the presence of SDS or high concentrations of urea. See, e.g., Short, et al., WO 02/44336.

Cannulae nanotubules are characteristically formed by *Pyrodictium abyssi*, a hyperthermophilic microorganism discovered in a high temperature environment (>100°C). In its natural environment and in cell culture, *Pyrodictium abyssi* are linked together by a meshwork of these nanotubular fibers that both connect and entrap the cells. These fiber networks are a unique feature of the genus *Pyrodictium* and they appear to be required for growth above 100°C. In addition, there appears to be a direct association between the maintenance of these nanotubular connections and cellular growth as demonstrated by the observation that, at the onset of cellular fission, these nanotubules appear to form loops attached at both ends to the growing cell. Following cellular fission the nanotubular loops become links connecting daughter cells. While it remains speculative as to what the true role of the nanotubules is in nature, it has been suggested that the linkage of cells by these tubules could enable cells to exchange metabolites, 10 genetic information, or signal compounds.

SUMMARY

The invention provides chimeric polypeptides comprising at least a first domain comprising a cannulae polypeptide and at least a second domain comprising a heterologous polypeptide or peptide. The heterologous polypeptide or peptide can be 20 inserted at the amino terminal end, the carboxy terminal end or internal to the cannulae polypeptide, or, if the cannulae polypeptide comprises more than one heterologous polypeptide or peptide, a mixture thereof. The cannulae polypeptide can comprise a protein having at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or more, sequence identity to SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, and 25 is capable of assembling into a polymer, e.g., a nanotubule, or, is capable of acting as a chiral selector. The chimeric cannulae proteins can assemble into nanotubular polymers to act as chiral selectors, biosynthetic pathways, selection scaffoldings and the like.

In one aspect, the cannulae polypeptide is capable of assembling into a polymer, such as a nanotubule. In one aspect, the cannulae polypeptide is capable of self-assembling into a polymer. In some aspects, the monomers require a co-factor for polymer assembly, e.g., a divalent cation, or, a "nucleation factor," which can be another cannulae monomer. The divalent cation can be Ca²⁺, Mg²⁺, Cu²⁺, Zn²⁺, Sr²⁺, Ni²⁺, Mn²⁺ and/or Fe²⁺. In another aspect, both Ca²⁺ and Mg²⁺ are needed for polymer assembly, e.g.,

into nanotubules. In one aspect, divalent cation(s) are in millimolar concentrations during polymer assembly.

In one aspect, the heterologous polypeptide or peptide is expressed in the inner lumen of a nanotubule or on the exterior of the nanotubule. These hybrid nanotubules can array the heterologous polypeptides or peptides on the outer surface or the inner luminal surface of a tubular polymer, or, when a monomer comprises more than one heterologous peptide or protein, they can be "displayed" on both the outer and inner surfaces of the tubules. If all the monomers of a nanotubule comprise a heterologous polypeptide or peptide in a similar manner, then that heterologous polypeptide or peptide can be displayed in a regular helical pattern on the nanotubule.

In one aspect, the heterologous polypeptide or peptide comprises a chiral selection motif, a receptor or a ligand, an enzyme, an enzyme active site, a cofactor, a substrate, an antigen or an antigen binding site, a detectable moiety, e.g., a green fluorescent protein, an alpha-galactosidase or a selection factor, e.g., a chloramphenicol acetyltransferase.

In one aspect, the chimeric polypeptide is a recombinant protein, which can be expressed *in vitro* or *in vivo*, a synthetic protein, or a mixture thereof.

In one aspect, at least one subsequence of the cannulae polypeptide domain of a chimeric protein of the invention has been removed. A heterologous polypeptide or peptide can be inserted into the cannulae polypeptide at the site (or one of the sites) subsequence(s) were removed. In one aspect, the cannulae polypeptide is a CanA polypeptide and the removed subsequence is a 14 residue motif (peptide) consisting of residue (position) 123 to residue 136 of SEQ ID NO:2 (i.e., "PDKTGYTNTSIWVP"), or, a 17 residue motif (peptide) located at amino acid residue (position) 123 to residue 139 of SEQ ID NO:2, (i.e., "PDKTGYTNTSIWVPGE"). The heterologous polypeptide or peptide can be inserted into the CanA polypeptide at one or both of the sites of the 14 or 17 residue motif subsequences were removed. The heterologous peptide can be a 14 residue or a 17 residue peptide inserted into the CanA polypeptide to replace the removed 14 residue or 17 residue motif.

The invention provides immobilized chimeric polypeptides comprising a chimeric monomeric or polymeric polypeptide of the invention. The invention provides polymers, e.g., nanotubules, comprising a plurality of chimeric polypeptides of the invention. In one aspect, the polymer is a heteropolymer, e.g., a nanotubule assembled from more than one cannulae polypeptide, including monomers other than the chimeric

proteins of the invention, or other polypeptides or compositions. The heterologous polypeptide or peptide comprises an enzyme, e.g., an active site, or a plurality of different enzymes. The plurality of enzymes can comprise a biosynthetic pathway. The plurality of enzymes can be arranged along the length of the nanotubule in the same order as they act in the biosynthetic pathway. The different enzymes comprising the biosynthetic pathway can be separated from each other along the length of the tubule by cannulae monomers lacking a heterologous protein or peptide (e.g., a “wild type” cannulae monomer, such as CanA, CanB, CanC, CanD and the like). The polymers comprising a biosynthetic pathway can also comprise substrate(s), co-factor(s), regulatory agents and the like.

10 The invention provides polymers, e.g., nanotubules, wherein the heterologous polypeptide or peptide comprises at least one chiral selection motif, such as an enzyme or an enzyme active site.

15 The invention provides nucleic acids comprising a sequence encoding a chimeric polypeptide of the invention. The invention provides expression cassettes (e.g., vectors, recombinant viruses, phages, etc.) comprising a sequence encoding a chimeric polypeptide of the invention. The invention provides cells comprising a sequence encoding a chimeric polypeptide of the invention, or, an expression cassette of the invention. The cell can be any cell, e.g., a bacterial cell, a plant cell, a yeast cell, a fungal 20 cell, an insect cell or a mammalian cell. The invention provides transgenic non-human animals comprising a sequence encoding a chimeric polypeptide of the invention, or, an expression cassette of the invention. The invention provides plants comprising a sequence encoding a chimeric polypeptide of the invention, or, an expression cassette of the invention.

25 The invention provides methods for the chiral selection of a composition, comprising the following steps: providing a chimeric polypeptide of the invention; providing a racemic mixture of the composition; and, contacting the racemic mixture with the chimeric polypeptide under conditions wherein only one enantiomer of the composition binds to the chimeric polypeptide; thereby selecting a single chiral specie of the racemic mixture. The invention provides methods for the chiral selection of a composition, comprising the following steps: providing a nanotubule of the invention; providing a racemic mixture of the composition; and, contacting the racemic mixture with the nanotubule under conditions wherein only one enantiomer of the composition binds to

the nanotubule; thereby selecting a single chiral specie of the racemic mixture. The methods further comprise separation of the different chiral species.

The invention provides methods for enzymatic biosynthesis of a composition, comprising the following steps: providing a nanotubule of the invention comprising a plurality of enzymes comprising a biosynthetic pathway; providing a substrate for at least one enzyme; and, contacting the nanotubule with the substrate under conditions wherein the enzymes of the biosynthetic pathway catalyze the synthesis of the composition. In one aspect, the enzymes are expressed in the inner lumen of the nanotubule, or, they are expressed on the exterior of the nanotubule. The nanotubules can also comprise substrates(s), co-factor(s), regulatory factors and the like.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, patents, patent applications, GenBank sequences and ATCC deposits, cited herein are hereby expressly incorporated by reference for all purposes.

DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Figure 1 is an illustration of a transmission electron micrograph of nanotubules assembled from recombinant CanA expressed in *E. coli*.

Figure 2 is a schematic representation of the open reading frames of the CanA and CanB sequences, showing the CanA sequence containing a 14 amino acid domain not found in CanB.

Figure 3 is an illustration of an immunofluorescent light microscope image of nanotubules assembled from a fusion protein generated by fusing the CanA open reading frame (SEQ ID NO:1) to the open reading frame of the green fluorescent protein ZSGREEN™.

Figure 4 is an illustration an exemplary process for constructing a heteropolymer of the invention generated by self-assembly of different chimeric monomers, as described below.

Like reference symbols in the various drawings indicate like elements.

5

DETAILED DESCRIPTION

The invention provides compositions (including chimeric proteins and nanotubules) and methods for use in all aspects of nanotechnology. The compositions of the invention, e.g., chimeric proteins and nanotubules, can be used in any biological or synthetic system. For example, the chimeric proteins and nanotubules of the invention can be used in electronic devices, such as circuits, transistors, memory storage devices or any aspect of a computer, transmitter, detector and the like. The chimeric proteins and nanotubules of the invention can be used in any pharmaceutical, medical device, artificial organ, prosthesis, implant, and the like, for example, as a structural element or coating. The chimeric proteins and nanotubules of the invention can be used in any article of manufacture, e.g., for biosynthetic scaffolding, camouflage or as heat resistance structural elements.

In one aspect, the chimeric proteins and nanotubules of the invention comprise a detectable moiety, e.g., a fluorescent protein, a radioactive moiety, an epitope and the like.

20

The invention provides compositions and methods for the identification, separation or synthesis of proteins or ligands using chimeric cannulae polypeptides (i.e., fusion, or hybrid, proteins). Chimeric cannulae polypeptides of the invention include CanA fusion proteins comprising SEQ ID NO:2 (encoded by SEQ ID NO:1), CanB fusion proteins comprising SEQ ID NO:4 (encoded by SEQ ID NO:3), CanC fusion proteins comprising SEQ ID NO:6 (encoded by SEQ ID NO:5), CanD fusion proteins comprising SEQ ID NO:8 (encoded by SEQ ID NO:7), or subsequences thereof.

25

In one aspect, the compositions and methods are used for the chiral separation of proteins and other compositions. For example, cannulae (e.g., CanA, CanB, CanC, CanD) fusion proteins can be used as chiral separations material. The chimeric cannulae polypeptides of the invention can be used as chiral separation materials in monomer or polymer (e.g., nanotubule) forms. When used in nanotubule forms, the motif of the cannulae polypeptide responsible for chiral selectivity can be exposed to the inner lumen of the tubule or on the outer surface of the tubule, or both.

The invention provides cannulae (e.g., CanA, CanB, CanC, CanD) fusion proteins comprising a cannulae polypeptide further comprising a heterologous polypeptide or peptide. The heterologous polypeptide or peptide can be an enzyme, an enzyme active site, a ligand, a receptor, an antigen, an epitope, or an antibody. The heterologous polypeptide or peptide can be any sequence for the chiral selection of a protein or other composition. For example, a chiral selection heterologous polypeptide or peptide can be an enzyme or an enzyme active site motif.

In one aspect, the cannulae fusion proteins are monomeric or polymeric, e.g., dimers, trimers, etc., or nanotubules, as illustrated in Figure 1. Cannulae chimeric polymers, e.g., nanotubules, can act as high density preparation materials, e.g., where the heterologous polypeptide or peptide comprises a chiral selection motif.

Cannulae chimeric polymers, e.g., nanotubules, also can act as a high density selection materials, e.g., where the heterologous polypeptide or peptide comprises a receptor, ligand, epitope, antibody and the like. In aspects where the cannulae chimeric polymers form as nanotubules, the heterologous polypeptide or peptide can be expressed on the outer surface of the nanotubule, on the inner surface of the tubule's lumen, or both.

Cannulae chimeric polymers, e.g., nanotubules, also can act as a biosynthetic scaffolding, e.g., where nanotubules of the invention comprise a plurality of heterologous polypeptide or peptides in the form of enzymes, catalytic antibodies or enzyme active sites comprising a biosynthetic pathway. In one aspect, the enzymes, catalytic antibodies or enzyme active sites are all expressed on one surface of a nanotubule, e.g., on the outer surface or on the inner lumen of the tubule. In one aspect, the enzymes, catalytic antibodies or enzyme active sites are arranged along the length of the tubule in the same order of their action in the biosynthetic pathway. Any number of enzymes, catalytic antibodies or enzyme active sites can be immobilized onto a tubule. Any biosynthetic pathway can be reconstructed along a nanotubule of the invention.

In one aspect, nanotubules comprising a plurality enzymes, catalytic antibodies and/or enzyme active sites are generated by constructing a cannulae polypeptide-enzyme fusion protein by fusing the open reading frame of a cannulae polypeptide (e.g., CanA, CanB) to the open reading frame of a desired enzyme sequence using standard molecular cloning techniques. The fusion sequence is then cloned into an appropriate expression cassette, e.g., an over-expression vector, prokaryotic or eukaryotic, and expressed as recombinant proteins. Expressed fusion protein can be purified from host proteins before polymer assembly. For example, chimeric proteins (e.g., monomers)

can be purified by heat treatment to denature heat-labile host proteins (e.g., at about 80 to 100°C, for about 2 to 20 minutes). The soluble heat-stable fusion protein can be further purified from contaminating proteins by other conventional means, e.g., chromatography techniques, e.g., ion exchange chromatography, HPLC and the like.

5 Purified, partially purified or unpurified chimeric (fusion) proteins can be induced to assemble into nanotubules by heating the fusion monomer solution (e.g., to about 80°C) in the presence of millimolar concentrations of a bivalent cation, e.g., calcium and/or magnesium. The polymer can be collected, e.g., by centrifugation (e.g., at 30,000 x g for 30 minutes), chromatography and the like.

10 The invention provides heteropolymers, e.g., nanotubules, comprising any variety of compositions, such as enzymes, catalytic antibodies and/or enzyme active sites, co-factors, substrates and the like to construct a biosynthetic pathway along the length of the polymer (e.g., nanotubule). Heteropolymers (e.g., nanotubules) of the invention can also comprise any variety of antibodies, antigens, receptors, ligands, binding sites and the 15 like, spatially arranged in any desired manner along the length of the polymer.

20 Heteropolymers (e.g., nanotubules comprising a plurality of different enzymes, catalytic antibodies and/or enzyme active sites comprising a biosynthetic pathway) can be constructed by an exemplary protocol as illustrated in Figure 4. Nucleic acids encoding chimeric monomers are constructed and expressed. The heterologous protein or peptide can be inserted at the amino terminal, carboxy terminal (as shown in Figure 4) or internal to the cannulae polypeptide (e.g., CanA). One or more, or all, or the 25 expressed chimeric monomers can be purified. Self-assembly of the heteropolymer can be initiated with one of the chimeric polypeptides, e.g., fusion 1 monomer pool as shown in Figure 4. Next, in this exemplary protocol, fusion 1 polymer is rapidly diluted with fusion 2-monomer pool such that the majority of the subunits added to the growing polymer are fusion 2 monomers. Alternatively, unassembled fusion 1 monomers are removed and fusion 2 monomers added. The resulting polymer is composed of a length of fusion 1 and a length of fusion 2 monomer. This process can be iteratively repeated until a nanotubule of a desired length comprising a desired number of different enzymes, 30 catalytic antibodies and/or enzyme active sites comprising a biosynthetic pathway is generated. The resulting nanotubule can serve as a scaffold for the assembly of an oriented, multi-enzyme complex.

In alternative aspects, the invention provides heteropolymers comprising different ratios of fusions and wild-type, non-fusion monomers to assemble nanotubular

polymers that display one or more enzyme (or other, e.g., binding or co-factor) activities, at controlled loading, on the exterior or interior surface of a nanotubule.

In one aspect, any number of compositions desired to be immobilized along the length of a polymer of the invention (e.g., a nanotubule), whether a protein or a 5 non-protein composition, e.g., enzymes, catalytic antibodies, enzyme active sites, co-factors (e.g., NADH, FADH, ATP and the like), substrates, antibodies, antigens, receptors, ligands, binding sites and the like, can be spatially arranged in any desired manner along the length of the polymer by indirect immobilization to the polymer. In this aspect, immobilization agents (e.g., receptors, ligands, antibodies, epitopes, substrates and the 10 like) are arranged as desired along the length of the polymer. The composition to be immobilized can be constructed to include (e.g., a chimeric recombinant protein) or be complexed with a moiety that will bind to an indirect immobilization agent. The indirect immobilization agent can be a binding agent for the composition to be immobilized. For example, a nanotubule is constructed having ten different antibodies spatially arranged 15 along the length of the tubule. This nanotubule can be constructed by a method analogous to that illustrated in Figure 4, e.g., instead of chimeric monomers comprising enzymes, the chimeric monomers would comprise antibodies (including, e.g., antigen binding sites) that specifically bind to different, desired enzymes, substrates, co-factors and the like.

20 In one aspect, the chimeric cannulae proteins of the invention self-assemble into helical nanotubular protein polymers. These helical nanotubular protein polymers can act as a chiral selectors, biosynthetic pathways, selection scaffoldings and the like. These hybrid protein nanotubules can array the heterologous polypeptide or peptide (fusion partner) on the outer surface or the inner luminal surface of a tubular 25 polymer. If all the monomers of a nanotubule comprise a heterologous polypeptide or peptide in a similar manner, then that heterologous polypeptide or peptide can be displayed in a regular helical pattern on the nanotubule.

In addition to serving as chiral selectors, biosynthetic pathways, selection scaffoldings, etc. comprising chimeric monomers of the invention, polymers of the 30 invention (e.g., nanotubular protein polymers) can also comprise unmodified cannulae monomers, modified non-chimeric cannulae monomers or other polypeptides. For example, in one aspect, a nanotubule of the invention comprises a chimeric monomer A, an unmodified cannulae monomer, a chimeric monomer B, etc. Inclusion of unmodified cannulae monomers can provide “spacing” between the “clusters” of heterologous

peptides or polypeptides expressed on the inner or outer surface of a nanotubules (spatially arranged, e.g., as illustrated in Figure 4). In one aspect, a polymer of the invention is designed to comprise a mix of proteins having different stabilities under different conditions, e.g., a nanotubule comprising temperature stable and temperature 5 labile monomers (chimeric or wild type, e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8). In one aspect, a polymer of the invention is designed to comprise a mix of different proteins, e.g., cannulae polypeptides, including chimeric, wild type or otherwise modified, e.g., non-thermostable.

10 In one aspect, a subsequence of a cannulae protein is removed and replaced by the heterologous polypeptide or peptide, or, the heterologous polypeptide or peptide can be added to a cannulae monomer. The removed subsequence can be amino- or carboxy-terminal, or, it can be internal to the cannulae protein. In one aspect, the removed subsequence is a motif that is expressed on the inner surface and/or the exterior surface of a cannulae nanotubule. Thus, when the removed sequence is expressed by a 15 heterologous sequence, the heterologous sequence is also expressed on the inner or the outer surface (or both) of the tubule.

20 In one aspect, for the fusion (hybrid) CanA protein, the removed subsequence consists of a 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 (i.e., "PDKTGYTNTSIWVP"), or, a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2, (i.e., "PDKTGYTNTSIWVPGE"). In one aspect, the removed sequence is replaced by a heterologous polypeptide or peptide. When the CanA monomer is in polymeric nanotubular form, a 14 residue motif consisting 25 of residue 123 to residue 136 of SEQ ID NO:2 (i.e., "PDKTGYTNTSIWVP"), or, a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2, (i.e., "PDKTGYTNTSIWVPGE") is expressed on the outer surface of the nanotubule. In this aspect, the CanA monomer protein can act as a chiral selector on the outer surface. If all the monomers of a nanotubule comprise a heterologous polypeptide or peptide inserted in or near this motif position (as an addition or a full or partial replacement for the CanA motif), then that heterologous polypeptide or peptide can be displayed in a regular helical 30 pattern on the outer surface of a CanA nanotubule. In one aspect, a 14 residue or a 17 residue heterologous peptide replaces the removed 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 or the 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2.

In one aspect, the chimeric cannulae protein of the invention, either in monomeric or polymer (e.g., nanotubule) form, are stable to a variety of conditions, e.g., temperature, pHs, chaotropic agents, detergents and the like. In one aspect, a polymer of the invention comprises is a heteropolymer comprising monomers of different stabilities under different conditions.

In one aspect, the monomers and polymers of the invention are used as chiral selectors, and methods for using these compositions for the chiral selection of compositions from racemic mixtures. The net charge and electrophoretic mobility of a protein chiral selector can be directly affected by the pH of the buffer solution (e.g., aqueous buffers) used during the separation. In one aspect, the separation methods of the invention (e.g., the chiral separation methods using cannulae fusion (hybrid) proteins as a chiral selectors) are practiced over a range of pH values. The pH of the buffer solution for use in the separations methods can be varied and optimal pH can be determined by routine screening. In one aspect, the methods are practiced over an operating range from about pH 5.5 to 8.5, or, pH 3 to pH 10, or, pH 2.5 to pH 11.

In one aspect, the separations methods of the invention (e.g., chiral selections) are practiced over a range of pH values and in the presence of SDS and/or urea. The presence of SDS and/or urea can improve aqueous chiral separations; see, e.g., Bojarski (1997) Electrophoresis 18:965-969. The stability screenings can be conducted as follows: purified recombinant cannulae monomer protein is assembled into polymer using an *in vitro* assembly protocol at neutral pH. Following completion of the assembly reaction, the sample is centrifuged and pelleted cannulae polymer collected.

The stability of nanotubules comprising cannulae fusion (hybrid) proteins can be affected by the buffer environment used in practicing the methods of the invention. The separation methods of the invention (e.g., the chiral separation methods) can be practiced in a variety of commonly used organic modifiers. In one aspect, organic modifiers are added to buffers used in practicing the methods of the invention to improve the resolution of enantiomers. The concentration of modifiers for use in the separations methods can be varied and optimal concentrations can be determined by routine screening.

In one aspect the invention provides methods to evaluate the stability of the polymers of the invention in the presence of commonly used organic modifiers, e.g., as listed in the following table:

Organic Modifier	Concentration Range
Methanol	0-15%
Ethanol	0-15%
1-propanol	0-15%
2-propanol	0-15%
acetonitrile	0-15%

These modifiers are organic modifiers commonly used in protein-based chiral selection methods development. The methods of the invention incorporate these and other organic modifiers and protocols as discussed by, e.g., Busch (1993) J. of Chromatography A. 635:119-126; De Lorenzi (1997) J. of Chromatography A. 790:47-64; Ahmed (1997) J. of Chromatography A. 766:237-244. All of the analytical methods used for the evaluation of polymer stability in aqueous buffers also may be compatible with buffers containing up to 15% (v/v) of these organic modifiers. The choice of buffer and buffer pH used for organic modifier screenings can incorporate the results of aqueous buffer stability studies. In one aspect, these modifiers are analyzed in buffers between pH 6.5 and pH 8.0, or, between pH 5.5 and pH 9.0, or between pH 4.5 and pH 10.0.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or polymers of the invention and the yield of the chiral selection methods of the invention are determined using capillary electrophoretic methods. In one aspect, the chiral selectivity method is evaluated using capillary electrophoretic methods and racemic mixtures of commercially available compositions, e.g., beta-blockers or equivalents.

These methods also can be used to evaluate the efficiency (e.g., the chiral selectivity) of various embodiments of the invention, e.g., regular, helical nanotubules comprising chimeric and/or wild type CanA, CanB, CanC, CanD, etc. or mixed species polymers.

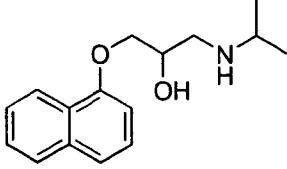
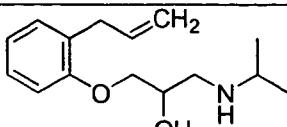
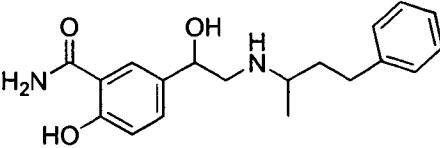
Data obtained from stability studies also can be used to determine by routine screening optimal buffer pH, acceptable additives, and organic modifier concentrations, depending on the desired outcome of a particular chiral separation protocol.

In one aspect, the resolution obtained with polymers (e.g. nanotubular chimeric cannulae) and monomers of the invention is determined using commercially available chiral selectors. There are numerous published methods for separating racemic mixtures of racemic compositions, e.g., beta-blockers, using commercial chiral selectors with, e.g., capillary electrophoresis. These methods can utilize both protein and non-protein chiral selectors. In one aspect, tests incorporating commercially available

enantio-separations media provide data about the comparative efficiency of nanotubular chimeric cannulae polymers and monomers of the invention as chiral selectors.

In one aspect, a chiral selectivity method of the invention or the resolution obtained with a polymer (e.g., nanotubular chimeric cannulae) and/or monomer of the invention is evaluated using capillary electrophoretic methods and racemic mixtures of commercially available beta-blockers, such as, e.g., those listed below:

Compound	Structure
Sotalol	
Atenolol	
Acebutolol	
Pindolol	
Metoprolol	

Propranolol	
Alprenolol	
Labetalol	

In one aspect, a chiral selectivity method of the invention or the resolution obtained with polymers and monomers of the invention is evaluated using capillary electrophoretic methods and racemic mixtures of propanolol. There are numerous reports in the literature that describe the resolution of enantiomers of propanolol, making it a good benchmark for the routine screening for optimizing chiral separations methods conditions employing the compositions of the invention, e.g., chimeric cannulae monomers and polymers (including nanotubules). Enantioseparation of propanolol has been accomplished using quail egg white riboflavin binding protein (see, e.g., De Lorenzi (1997) *supra*), pepsin, cellobiohydrolase, and bovine serum albumin (see, e.g., Tanaka (2001) *J. of Biochem. Biophysical Methods* 48:103-116; Henriksson (1996) *FEBS Letters* 390:339-344).

In one aspect, monomers of polymers of the invention are immobilized on a surface, e.g., a capillary. In one aspect, the methods of the invention are practice in a capillary tube, e.g., a GIGAMATRIX™ (Diversa Corporation, San Diego, CA). Both untreated and polyacrylamide-coated capillaries can be used to practice the methods of the invention. Untreated capillaries may be unsuitable for chiral selection due to adsorption of a chiral selector or an analyte on the walls of the capillary, see, e.g., Tanaka (2001) *supra*.

As discussed above, any separation fluid or organic modifier can be used to practice the methods of the invention. Determining optimal conditions by routine screening can be based on an optimization procedure described by Allenmark, S.G.

Chromatographic Enantioseparation. Methods and applications. pg 90-141. 1998. West Sussex, England, Ellis Horwood Limited. This exemplary protocol uses a neutral buffer without additives or modifiers as the starting condition for separation. If the enantiomers are not resolved, the pH can be adjusted to pH 5.5 or 8.5. If one of these pH conditions results in loss of sample due to excessive complexation with a chimeric cannulae monomer or polymer of the invention, a low percentage of an organic modifier can be introduced. Changes also can be made to the buffer pH, choice of organic modifier, and concentration of organic modifier to improve resolution.

In one aspect, routine screening methods are carried out using a partial filling technique, as described, e.g., by Tanaka (2001) *supra*; Chankvetadze (2001) *J. of Chromatography A*. 906:309-363. In this exemplary technique the capillary (e.g., GIGAMATRIX™, Diversa Corporation, San Diego, CA) is only partially filled with the protein chiral selector (a chimeric cannulae monomer or polymer of the invention). This can minimize the sensitivity issues associated with the high UV backgrounds produced by protein at the detector. Using this method, it is possible to use up to 500 uM protein during the enantioseparation.

A countercurrent technique can also be used. In countercurrent separations, conditions are used such that there is electrophoretic migration of the protein chiral selector (a chimeric cannulae monomer or polymer of the invention) away from the detector while the analyte migrates past the detector, see, e.g., Chankvetadze (2001) *supra*.

In alternative aspects, monomer or polymers or mixtures thereof are used to practice the methods of the invention. Chimeric cannulae monomers can have the ability to self-assemble into nanotubules. In one aspect, the chiral resolving power of different polymers (e.g., heteropolymers comprising chimeric and wild type cannulae proteins) and monomers relative to the resolving power of other polymers and monomers can be determined by routine screening, e.g., as described herein. By assembling into a nanotubule, a chimeric cannulae protein becomes a macromolecular structure that possesses distinct microenvironments, including an interior surface, cavity and an exterior surface. In addition, the regular assembly of the subunits into a helical structure introduces additional chirality into the polymer. The polymers of the invention include varying amounts of chirality, as varying amounts of chirality can enhance the enantioselectivity of the composition. The monomers and polymers of the invention can

be designed to have varying constrained quaternary (4⁰) structures. In one aspect, varying constrained quaternary (4⁰) structures results in varying amounts of chiral selectivity.

5 In one aspect, the chiral selection is performed under cooling conditions and in the absence of sufficient divalent cation (less than 1 mM) so a cannulae monomer (e.g., a CanA monomer) will not self-assemble during chromatography.

In one aspect, the performance of the chiral selective compositions of the invention are compared to the performance of commercially available chiral selectors. In one aspect, beta-blocker resolutions are performed with capillaries packed with 10 cellobiohydrolase or α_1 -acid glycoprotein (ChromTech AB Cheshire, UK) using, e.g., the separation conditions provided by the supplier. Comparisons also can be made to separations obtained using highly sulfated cyclodextrins (Beckman Coulter, Fullerton, 15 CA) according to, e.g., methods available from their applications guide. Other characteristics, such as good stability or minimal interference with analyte detection, can also be evaluated. Chimeric cannulae proteins of the invention, including the chimeric CanA polypeptide made by inserting peptide domains into a nonessential surface-exposed domain of CanA (see Figure 1), can be evaluated using these routine screening methods. Figure 1 is an illustration of a transmission electron micrograph of nanotubules assembled 20 from recombinant CanA expressed in *E. coli*.

Because of the macromolecular similarity of nanotubular cannulae 25 polymers to eukaryotic microtubules, any of the analytical methods that have been established in the microtubule field can be used to analyze chimeric cannulae polymers of the invention; see, e.g., Frederiksen, D.W. and L.W. Cunningham. Structural and Contractile Proteins, Part B: The Contractile Apparatus and the Cytoskeleton. 1982. Methods in Enzymology 85[Part B]. In evaluating the chiral selectivity of a chimeric cannulae monomer and/or polymer of the invention and the yield of the chiral selection 30 methods of the invention, light and electron microscopy (e.g., transmission electron microscopes), differential centrifugation, size exclusion chromatography and/or turbidity measurement methods can be used. Each of these methods provides slightly different information about the stability and integrity of the assembled chimeric cannulae polymer.

The assembly and disassembly of polymers of the invention can be followed by measuring changes in solution turbidity, e.g., as described in Purich, D.L., et al. (1982) Microtubule disassembly: a quantitative kinetic approach for defining endwise linear depolymerization. Methods in Enzymology 85[Part B], 439-450. In one aspect, kinetic turbidity measurements are used. Kinetic turbidity measurements can be used to

reflect changes in polymer weight concentration. These measurements can be used to determine rates of depolymerization. In one aspect, solution turbidity is monitored spectro-photometrically at 350 nm in a long path length cuvette. The long path length can provide an enhancement of the absorbance change improving sensitivity of the assay. In 5 one aspect, the method comprises a long path length and a temperature-controlled cuvette containing buffers that can range in pH from 3 to 10. Stability of polymer can be measured by diluting concentrated solutions of polymer into the cuvette containing temperature-equilibrated buffer.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or 10 polymers of the invention and the yield of the chiral selection methods of the invention are determined using a chiller-cooled system. The stability of polymer can be evaluated over a range of temperatures, e.g., from about 4°C to 80°C for each buffer pH. In one aspect, if it is not possible to measure accurate depolymerization rates using the rapid dilution method (due to over-dilution of the polymer into the test buffer), a resuspension 15 method can be utilized. In the resuspension method, a wide-bore pipette can be used to resuspend polymer pellets in temperature-equilibrated buffer. The resuspended pellet then can be transferred to a cuvette for analysis. The advantage of this method is the ability to use more concentrated polymer solutions. The drawback, however, is variability introduced by potential shearing of the polymer during resuspension.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or 20 polymers of the invention and the yield of the chiral selection methods of the invention are determined using differential centrifugation. Differential centrifugation can be used to assess the distribution of monomer protein incorporated into polymer vs. monomer free in solution. The differential centrifugation assay is useful for longer time course stability 25 evaluations. In these assays, polymer that has been assembled under standard conditions at neutral pH can be pelleted by centrifugation and then resuspended in a buffer (e.g., at varied pH, such as from pH 3 to pH 10) and pre-equilibrated at a specified temperature (e.g., at varied temperature, such as a range from about 4°C to 80°C). The samples can be incubated at temperature for 2 to 24 hrs and then re-centrifuged to pellet the intact 30 polymer. The supernatant and pellet fractions can be analyzed by SDS-PAGE. The supernatant will contain any soluble monomer (released by polymer depolymerization) and the pellet will contain intact polymer.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or polymers of the invention and the yield of the chiral selection methods of the invention

are determined using size exclusion chromatography. Size exclusion chromatography can be used to analyze the overall size distribution of polymers. Polymer samples can be resuspended in buffer (e.g., at varied pH, such as from pH 3 to pH 10) and incubated for 24 hours at 4°C. Following incubation, the samples can be fractionated, e.g., on a

5 Sephadryl S-1000 column (Amersham Pharmacia, Piscataway, NJ). This size exclusion column will separate the micron length polymer from shorter polymers and oligomers. Because polymers can be extremely stable at 4°C and neutral pH, and this buffer treatment can be used as the control.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or polymers of the invention and the yield of the chiral selection methods of the invention are determined using light microscopy, e.g., video-enhanced light microscopy, including both phase and differential interference contrast (DIC) optics. Light microscopy can be used to evaluate the gross morphology of polymers following extended incubations (e.g., between about 24 to 48 hours) at varied pH, such as from pH 3 to pH 10. Light microscopy can provide useful information about the extent of nanotubule polymer bundling. It also can be used to detect the presence of larger protein aggregates.

In one aspect, the chiral selectivity of chimeric cannulae monomer and/or polymers of the invention and the yield of the chiral selection methods of the invention are determined using electron microscopy (EM), e.g., standard negative stain transmission electron microscopy. Electron microscopy can be used to look at the fine structure of nanotubules. EM can be useful for the analysis of periodicity and helicity of the intact polymers. In addition, EM can detect other protein assemblies that may form during incubation at various pH values or in the presence of organic modifiers. Depending on the incubation conditions, eukaryotic microtubules have been shown to assemble into a number of macromolecular structures including ring, sheets, and ribbons, as described, e.g., in Hyams, J.S. and C.W. Lloyd. *Microtubules*. 1993. New York, Wiley-Liss. The polymers of the invention can be modified to assemble or reassemble into such alternate structures.

Chimeric cannulae protein of the invention can be abundantly and economically expressed as a recombinant protein *in vitro* or *in vivo* using any expression system, including bacteria, yeast, mammalian or plant expression systems (e.g., as host cells), as discussed below.

In some aspects, the protein nanotubes of the invention have advantages over traditional carbon nanotubes (which were discovered by the Japanese microscopist

Sumio Iijima in 1991). These advantages can stem from the powerful capabilities afforded by a biological system over the carbon based system. In particular, the protein polymer of the invention is evolvable, allowing amino acid changes to be incorporated into the protein while preserving the polymer structure. In one aspect, the invention 5 provides methods for modifying polypeptides of the invention, as described herein. This allows the protein to be modified in a way that changes its chemical characteristics, such as charge and hydrophobicity, thereby greatly expanding the number of applications for the polymer. For example, compositions of the invention can comprise circuits and transistors.

10 With the ability of the biological nanotube proteins to bind to GFP proteins, nanotubes of the invention can be used as fluorescent tubes. In one aspect, this can greatly increase the accuracy of data received through microscopic viewing. Due to the ability of the biological nanotubes to modify their chemical makeup, there are endless medical applications for use of nanotubes of the invention. Presently, patients are 15 frequently rejecting transplants, and transfusions. However, in some aspects, with the use of biological nanotubes of the invention, the make-up of the proteins might be altered, and even be made into working organs and eliminating some problems with organ and blood rejection. On an electrical scale, due to the biological nanotubes' ability to change chemical make-up, the charge of nanotubes of the invention can also be manipulated to 20 become an excellent composition for conducting wires. Thus, nanotubes of the invention can be made into computer transistors, e.g., supercomputers, but at a minuscule size, e.g., in some aspects, only a few nanometers in length.

Unlike carbon nanotubes, biological nanotubes of the invention can be manufactured in microbes or in plants; in some aspects, at an exceptionally low cost. 25 While carbon nanotubes are manufactured by use of natural gases such as methane, the only necessary energy needed in manufacturing carbon nanotubes would be the light of the sun and water, used in plants. With this free source of energy, and no harmful byproducts in its production of these natural polymers, all that remains are the benefits of these biological nanotubes. Presently, carbon nanotubes are produced, with yield as low 30 as one to two pounds a day, costing up to nine hundred dollars a gram. With this incredibly high cost of manufacturing, it is currently unrealistic to expect research in the carbon nanotubes to be useful for all applications. In contrast, in some aspects, proteins of the invention can be generated at an exceedingly lower price, and with many applications.

Another important aspect of biological nanotubes of the invention is that unlike carbon nanotubes, the proteins of the invention can be evolved into continuously improved types of protein nanotubes for almost any need. Through the use of biological mutations, the protein can be manipulated to become a universal enzyme, capable of 5 manipulating itself into any substrate. In one aspect, it can be used for advanced camouflage in the military, e.g., in one aspect, it is able to change its color in any surrounding to match identically to the environment. Carbon nanotubes, though very strong, are rigid and do not form the same flexible shapes that proteins can. In some 10 aspects, protein nanotubes of the invention can be formed into any shape, can withstand temperatures up to 150°C, can change its chemistry to be a universal donor, and can be evolved into new proteins.

In some aspects, there are differences between carbon nanotubes and biological nanotubes of the invention. In one aspect, the first difference is that a nanotube of the invention can be generated from a living organism (note: in an alternative aspect, a 15 process of the invention can generate a nanotubules in a cell-free or synthetic system, e.g., *in vitro*). Secondly, the bonds are different in each case. In the carbon nanotubes, the carbon atoms are held tightly together by strong covalent bonds. Carbon nanotubes, also known as “bucky-balls” are composed of C60, sixty covalently bonded carbon atoms. Covalent bonds occur when atoms share electrons, either in “free-loader” bonds, single 20 bonds, double bonds, or triple bonds. Covalent bonds are the strongest type of molecular bonds, having high boiling points, but are also very rigid. Since carbon nanotubes contain rigid covalent bonds, they do serve as excellent materials for buildings, however, for more practical uses, their rigidness inhibits them from being used easily in everyday products, such as clothing, and other textiles. However, in the biological protein 25 nanotubes of the invention the proteins are held together by peptide bonds. These peptide covalent bonds are not as strong as the covalent bonds holding together the carbon in the carbon nanotubes. This may cause a lower heat resistance. However, with a heat resistance of up to approximately 150°C, protein nanotubes of the invention can still be used for any heat-resistant application, e.g., in earthquake resistant building materials and 30 other such applications. Also, with the peptide bonds, the tubules of the invention are more flexible, allowing them to be used in clothing or other such products without the restrictions of the carbon nanotubes. Thus, the nanotubules and processes of the invention can be used in any aspect of nanotechnology.

Cannulae Polypeptide and Peptides

The invention provides chimeric polypeptides comprising at least a first domain comprising a cannulae polypeptide and at least a second domain comprising a heterologous polypeptide or peptide. The chimeric (fusion) cannulae polypeptides of the invention can be recombinant proteins encoded by nucleic acids comprising fusion of the sequence of a cannulae monomer to other protein or peptide coding sequences (heterologous sequences) to produce cannulae fusion (chimeric) proteins. However, the chimeric (fusion) cannulae polypeptides of the invention can be joined to the heterologous polypeptide or peptide by any means, including linkers. The chimeric (fusion) cannulae polypeptides of the invention can be partly or entirely synthetic. In one aspect, the chimeric monomers of the invention can form dimers, trimers (polymers of any length) and/or they can assemble, e.g., self-assemble, into a higher order structure, e.g., a quaternary structure, such as a nanotubule. The heterologous sequences can be added to the cannulae protein's amino- or carboxy-terminal end, or, they can be added 10 internal to the cannulae protein.

In one aspect, a subsequence of a chimeric (fusion) cannulae polypeptide of the invention is removed. In one aspect, a subsequence of a chimeric (fusion) cannulae polypeptide of the invention is removed and replaced by a heterologous polypeptide or peptide. Alternatively, the heterologous polypeptide or peptide can be added to another 20 section of the monomer (i.e., distal to the removed subsequence). The removed subsequence can be amino- or carboxy-terminal, or, it can be internal to the cannulae protein. In one aspect, the subsequence of fusion (hybrid) CanA protein that is removed and replaced by a heterologous polypeptide or peptide is a 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 (i.e., "PDKTGYTNTSIWVP"), or, a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2, (i.e., "PDKTGYTNTSIWVPGE"). When the CanA monomer is in polymeric nanotubular 25 form, a 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 or a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2 is expressed on the outer surface of the nanotubule. In one aspect, the tubule can then act as a high-density chiral selector. The surface-exposed 14 or 17 amino acid domain in CanA 30 is not essential for self-assembly of nanotubules. Thus, these domains can serve as a site for the insertion of peptides, e.g., with chiral selector properties, ligand binding properties, and the like.

Once assembled, e.g., as a nanotubule, chimeric cannulae proteins of the invention can serve as a molecular scaffold that displays its heterologous sequence (its chimeric/fusion protein partner) in a defined orientation in a regular, helical array. This functional flexibility offers the opportunity to display a large variety of recombinant 5 proteins on the surface of a nanotubule to create chiral selectors with a wide range of applications. The heterologous sequences can be chiral selection motifs, enzymes, active sites, epitopes, ligands, receptors, antigens, antibodies or antigen binding sites, nucleic acid binding proteins, and the like.

In one aspect, the chimeric cannulae monomers are overexpressed in a host 10 cell, e.g., a bacteria such as an *E. coli*. In one aspect, the overexpressed polypeptide is modified by nucleic acid mutagenesis and/or directed protein evolution, as described herein.

The cannulae domain of the chimeric polypeptides of the invention can comprise a CanA polypeptide as set forth in SEQ ID NO:2 (encoded by SEQ ID NO:1); a 15 CanB polypeptide as set forth in SEQ ID NO:4 (encoded by SEQ ID NO:3); a CanC polypeptide as set forth in SEQ ID NO:6 (encoded by SEQ ID NO:5); a CanD polypeptide as set forth in SEQ ID NO:8 (encoded by SEQ ID NO:7); a CanE polypeptide as set forth in SEQ ID NO:10 (encoded by SEQ ID NO:9). The cannulae domain of the chimeric polypeptides of the invention also can comprise a polypeptide 20 having a 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to polypeptide as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, wherein the 25 cannulae domain polypeptide can form a nanotubule and/or can act as a chiral selector (in monomeric or polymeric form). The cannulae domain of the chimeric polypeptides of the invention also can comprise a polypeptide encoded by a nucleic acid having a 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a nucleic acid as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, wherein the cannulae domain polypeptide can form a nanotubule and/or can act as a chiral selector (in monomeric or polymeric form). The cannulae domains of the chimeric polypeptides of the invention 30

can comprise two or more of these proteins, including mixtures of CanA, CanB, CanC, CanD and/or CanE.

The term protein or polypeptide sequence or amino acid sequence includes an oligopeptide, peptide, polypeptide, or protein sequence, or to a fragment, portion, or 5 subunit of any of these, and to naturally occurring or synthetic molecules. The terms “polypeptide” and “protein” include amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain modified amino acids other than the 20 gene-encoded amino acids. The term “polypeptide” also includes peptides and polypeptide fragments, motifs and the like. The term also includes 10 glycosylated polypeptides. The peptides and polypeptides of the invention also include all “mimetic” and “peptidomimetic” forms.

The invention also comprises “variants” of the chimeric polynucleotides or polypeptides of the invention, and methods of making them, wherein the variants are modified at one or more base pairs, codons, introns, exons, or amino acid residues 15 (respectively) yet retain the activity or have a modified activity of a chimeric polypeptide of the invention. Variants can be produced by any number of means included methods such as, for example, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific 20 mutagenesis, gene reassembly, GSSM and any combination thereof. Techniques for producing variant chimeric polypeptides having activity at a pH or temperature, for example, that is different from a template chimeric polypeptide, are included herein. The term “saturation mutagenesis” or “GSSM” includes a method that uses degenerate oligonucleotide primers to introduce point mutations into a polynucleotide, as described 25 in detail, below. The term “optimized directed evolution system” or “optimized directed evolution” includes a method for reassembling fragments of related nucleic acid sequences, e.g., related genes, and explained in detail, below. The term “synthetic ligation reassembly” or “SLR” includes a method of ligating oligonucleotide fragments in a non-stochastic fashion, and explained in detail, below.

30 In one aspect, nucleic acids encoding the chimeric polypeptides of the invention are cloned and over-expressed in a host cell, e.g., bacterial (e.g., *E. coli*, *Bacillus*, *Streptomyces*), yeast, plant or mammalian host cells.

Purified recombinant chimeric cannulae protein of the invention can self-assemble into nanotubules. In some aspects, presence of a divalent cation may be needed,

depending on the conditions and mixture of polypeptides comprising the nanotubular assembly or the presence of proteins that catalyze or facilitate tubule assembly. Thus, in one aspect the chimeric cannulae proteins of the invention or the nanotubules of the invention are assembled in the presence of a divalent cation. The divalent cation may be 5 Ca^{2+} , Mg^{2+} , Cu^{2+} , Zn^{2+} , Sr^{2+} , Ni^{2+} , Mn^{2+} and/or Fe^{2+} . In one aspect, a single divalent cation is needed, e.g., Ca^{2+} or Mg^{2+} . In another aspect, both Ca^{2+} and Mg^{2+} are needed for chimeric cannulae protein can self-assemble into nanotubules. In one aspect, the divalent cation(s) are present in millimolar concentrations.

In alternative aspects, chimeric proteins of the invention or nanotubules of 10 the invention are assembled in the presence of one or more initiators, which can be one or more environmental conditions, e.g., increased temperature, pH or salinity, and/or one or more compositions as an initiator, e.g., a partially polymerized monomer as a primer or any element found in the original environment of the *Pyrodictium abyssi* organism, from which the *canA* gene was initially derived. For example, in one aspect, the chimeric 15 proteins of the invention or nanotubules of the invention are assembled in the presence of seawater, or equivalent, from the growth microenvironment of the *Pyrodictium abyssi* organism, or equivalent organisms which form CanA-like nanotubules. For example, in one aspect, the chimeric proteins of the invention or nanotubules of the invention are assembled in the presence of black-smoker fluid, or equivalent. Equivalent environments 20 that can be used in the methods of the invention for the assembly of chimeric proteins of the invention or nanotubules of the invention include fluids comprising the same, substantially the same, or a selected subset of elements found in the growth microenvironment, e.g., black-smoker fluid, or equivalent. For example, the methods of the invention (e.g., for polymerizing chimeric polypeptides of the invention, or to 25 assemble nanotubules of the invention) can comprise use of any mixture of salts, e.g., iron sulfate, manganese sulfate, lead sulfate, lithium sulfate, manganese chloride and/or calcium chloride or equivalent salts. The invention provides methods for the controlled polymerization of proteins of the invention in the presence of different catalyst salts, such as iron sulfate, manganese sulfate, lead sulfate, lithium sulfate, manganese chloride 30 and/or calcium chloride or equivalent salts. In one aspect, the polymerization takes place in a solution. In one aspect, the controlled polymerization conditions can further comprise modification of temperature, salinity, pH and the like. The methods of the invention also can comprise use of some, or all elements described in Table 1 (e.g., H_2S , H_2 , CH_4 , Mn , Fe , Be , Zn , Cu , Ag , Pb , Co , Si , Al , Ba , Cs , Li , Rb , CO_2 , Ca , Sr , B , As Se , P ,

Mg, SO₄, and/or Alk), wherein the concentrations of elements set forth in Table 1 are only alternative embodiments to practice the assembly processes of the invention. In one aspect, copper sulfate salt is used as an initiation inhibitor or depolymerization element, particularly when used as an isolated element, versus one of many elements in a complex growth environment solution comprising many salts and elements.

The chimeric polypeptide of the invention can comprise the cannulae polypeptides CanA, CanB, CanC, CanD and/or CanE, and subsequences and mixtures thereof. In the following alignment, CanA and CanA_pep stand for nucleic acid SEQ ID NO:1 and its corresponding amino acid SEQ ID NO:2, respectively; CanB and CanB_pep stand for nucleic acid SEQ ID NO:3 and its corresponding amino acid SEQ ID NO:4, respectively; CanC and CanC_pep stand for nucleic acid SEQ ID NO:5 and its corresponding amino acid SEQ ID NO:6, respectively; CanD_partial stands for nucleic acid SEQ ID NO:7 or its corresponding amino acid SEQ ID NO:8; and CanE_partial stands for nucleic acid SEQ ID NO:9 or its corresponding amino acid SEQ ID NO:10.

15 Nucleic acid alignment for SEQ ID NOS:1, 3, 5, 7, and 9:

	1	50
	CanA (1) <u>GTC</u> AGTACACAA <u>ACC</u> T <u>AGC</u> CT <u>AT</u> A <u>GG</u> C <u>GT</u> A <u>TT</u> A <u>T</u> <u>CC</u> C <u>TC</u> CC <u>CT</u> CCCC <u>C</u>	
	CanB (1) <u>GTC</u> AG <u>CC</u> CT <u>AC</u> GG <u>GT</u> C <u>T</u> <u>AG</u> CC <u>CT</u> <u>CC</u> C <u>GT</u> <u>CT</u> <u>AT</u> <u>CA</u> T <u>T</u> <u>CC</u> C <u>TC</u> CC <u>CT</u> CCCC <u>G</u>	
	CanC (1) <u>A</u> <u>T</u> <u>G</u> <u>AG</u> <u>G</u> TAC <u>AC</u> GG <u>AC</u> <u>CC</u> T <u>AG</u> CT <u>CT</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CA</u> <u>TA</u> <u>AG</u> T <u>TC</u> CC <u>CT</u> CCCC <u>G</u>	
20	CanD_partial (1) -----	
	CanE_partial (1) -----	
	Consensus (1) TGA G AC C CTAGC T GC GG AT T GCCTCGGCTGCCG	
	51	100
	CanA (51) <u>C</u> <u>T</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>C</u> <u>T</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CT</u> <u>T</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>AC</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>CA</u> <u>AC</u> <u>CC</u> <u>T</u>	
25	CanB (51) <u>C</u> <u>T</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>C</u> <u>T</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CT</u> <u>T</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>AC</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>CA</u> <u>AC</u> <u>CC</u> <u>T</u>	
	CanC (51) <u>C</u> <u>T</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>C</u> <u>T</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CT</u> <u>T</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>AC</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>CA</u> <u>AC</u> <u>CC</u> <u>T</u>	
	CanD_partial (1) -----	ACCT
	CanE_partial (1) -----	ACCT
	Consensus (51) CCTCGCCCT CTAGCAGGCTTCGCCAC ACCCAGAGCCC CT A CAGCT	
30	101	150
	CanA (101) <u>T</u> <u>C</u> <u>T</u> <u>AC</u> <u>GG</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>G</u> TAC <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CA</u> <u>GT</u> A <u>AG</u> AC <u>GT</u> T <u>AC</u> <u>AA</u>	
	CanB (101) <u>T</u> <u>C</u> <u>T</u> <u>AC</u> <u>GG</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>G</u> C <u>CC</u> <u>CA</u> <u>A</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>AA</u> <u>T</u> <u>A</u> <u>CA</u> <u>CC</u> <u>GT</u> A <u>AG</u>	
	CanC (101) <u>T</u> <u>C</u> <u>T</u> <u>AC</u> <u>GG</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>G</u> C <u>CC</u> <u>CA</u> <u>A</u> <u>AC</u> <u>CT</u> A <u>AG</u> CC <u>AC</u> <u>CC</u> <u>AA</u> <u>T</u> <u>A</u> <u>CA</u> <u>CC</u> <u>GT</u> A <u>AG</u>	
35	CanD_partial (5) <u>T</u> <u>C</u> <u>T</u> <u>AC</u> <u>GG</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>G</u> C <u>CC</u> <u>CA</u> <u>A</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CA</u> <u>GT</u> A <u>AG</u> AC <u>GT</u> T <u>CT</u> <u>A</u>	
	CanE_partial (5) <u>T</u> <u>C</u> <u>T</u> <u>AC</u> <u>GG</u> <u>CC</u> <u>AC</u> <u>CC</u> <u>G</u> C <u>CC</u> <u>CA</u> <u>A</u> <u>AC</u> <u>CC</u> <u>AC</u> <u>GG</u> <u>CA</u> <u>GT</u> T <u>CT</u> <u>A</u>	
	Consensus (101) TCTACGCCACCGGCACAGCACAGGCAGTAAGCGAGCCAATAGACGTAGAA	
	151	200
	CanA (151) <u>A</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>CT</u> --- <u>CC</u> <u>C</u> <u>A</u> <u>G</u> <u>CA</u> <u>AA</u> <u>A</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CA</u> <u>A</u> <u>CC</u> <u>CC</u> <u>CA</u> <u>G</u> <u>CT</u> <u>CA</u>	
	CanB (151) <u>A</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>CT</u> --- <u>CA</u> <u>C</u> <u>A</u> <u>G</u> <u>CA</u> <u>AA</u> <u>A</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CA</u> <u>A</u> <u>CC</u> <u>CC</u> <u>CA</u> <u>G</u> <u>CT</u> <u>CA</u>	
40	CanC (151) <u>A</u> <u>CC</u> <u>CC</u> <u>AC</u> <u>CT</u> <u>TA</u> <u>G</u> <u>A</u> <u>CA</u> <u>A</u> <u>CC</u> <u>CA</u> <u>T</u> <u>A</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CT</u> <u>CC</u> <u>T</u> <u>CC</u> <u>CC</u> <u>GT</u> <u>CA</u> <u>AA</u>	

CanD_partial	(55)	ACGAGCCTCGGTACG---CTAAATACTGCCGTGGTGCACAGGGTAAGCA	
CanE_partial	(55)	ACGAAACCTTAACACGGCCATAGCCCCTGCTGCCGGCGCCAGGGCAGCGT	
Consensus	(151)	AGCCACCT CA CA CATAGCCCCTGCTGCCGGCGCACAGGGCAGC A	
	201		250
5			
CanA	(198)	CGACATAGGTACGCAATAGGTGCGATAAGGACCAAGGTCATGATGAA	
CanB	(198)	CGACATAGGTACTTCAACGCGACGCCAAGGATCAAGTGAACGTGACAA	
CanC	(201)	GGACATGGCTACATTAGATAACTAACAGTCAAAGTTAATGTAATAA	
CanD_partial	(102)	GACGCTAGGAGACATAACATAATATGCCGACATGACGTGAACATAACAA	
CanE_partial	(105)	GGGCATAGGAGCATAAACATAAGACAAACAGACTGACGTGAACGTTGTGA	
10			
Consensus	(201)	GGACATAGGCTACATAA AATA A CAAG AT A GTGAACGT ATAA	
	251		300
CanA	(248)	AGCTGAAGGTGACCCCTGGCTAACGCTGAGCAGCTAAAGCCCTACTTCAAG	
CanB	(248)	AGATAAAGGTGACCCCTGGCTAACGCTGAGCAGCTAAAGCCCTACTTCAAG	
CanC	(251)	AGCTGAAGGTGACTCTCGCTAACGGCGACAGCTAAAGCCCTACTTCAAG	
15			
CanD_partial	(152)	AGCTAACGCTACGGCTTCTAACGCTCACAGCTAAAGCCCTACTTCAAG	
CanE_partial	(155)	AGCTGAAGATAACCCCTGGCAACGCTGAGCAGCTAAAGCCCTACTTCAAG	
Consensus	(251)	AGCTGAAGGTGACCCCT GCTAACGCTGAGCAGCTAAAGCCCTACTTCAAG	
	301		350
20			
CanA	(298)	TACCTACAGATAACAGATAACAAGGGCTATGAGACGAACAGCACAGGTCT	
CanB	(298)	TACCTACAGATAAGTGTCTAAAGAGCG-----	
CanC	(301)	TACCTACAGCTAGTACACAAAGAAC-----GCCAC	
CanD_partial	(202)	TACCTGATAATAAGCTAGTAACGCT-----GGACAGG AA	
CanE_partial	(205)	TACCTACAGATAAGTGTCTAAAGAGCGT-----TGACAGG AA	
Consensus	(301)	TACCTACAGATAAGTGTCTAA AAGCG	ACAGC A
25	351		400
CanA	(348)	AGGAAATTCAAGCGAGACCAAGGCTGTGATAAGCTCGAACACCCAGCG	
CanB	(323)	ACGTAGCTGA--CGAGATCAAGGCCATAAGGATAGACAAGCCTAGCG	
CanC	(333)	TCCCACGGACA---TGGTAAAGGCTGCTAACGCTCGAGAAGCCTAGCG	
CanD_partial	(237)	CCCCAACGAGTCGGAGGAAAGGGCATGATAACTCTATGGAAGCCCTACG	
30			
CanE_partial	(240)	CGAGATCAAGGCTG-----TGCTAACGCTCGAGAAGCCAGCG	
Consensus	(351)	GGCA C A CGAG AAGGC GTGATAAGCCTCGAGAAGCCTAGCG	
	401		450
35			
CanA	(398)	CCCTGATACTAGTACAGACAAAGCAGCTATAGCAGTGTCTATCCGGACAAG	
CanB	(371)	CCGTCTATAATACTAGACAGCGAGGA-----	
CanC	(380)	CACTCTATAATACTAGACAAACGATGA-----	
CanD_partial	(287)	CCCTGATAATAGACACATCAAC-----	
CanE_partial	(278)	CACTCTATAATACTGGACAACGAGGA-----	
Consensus	(401)	CCGTCTATAATACTAGACAAACGAGGA	
	451		500
40			
CanA	(448)	ACCGGTTACACAAACACTTCGAATGCGTACCCGGTGAACCTGACAAGAT	
canB	(396)	-----CTTCGAC-----G-----	
CanC	(405)	-----CTACGATA-----G-----	
canD_partial	(312)	-----TTCTACACGACA-----	
canE_partial	(303)	-----CTTCGAGGGCGC-----	
45			
Consensus	(451)	CTTCGA A G	
	501		550

	CanA	(498)	AATTGT <u>CTACAA</u> CCGACAAAGCC <u>ACTAC</u> CTATA <u>TGAA</u> CTTCAAG <u>CCCT</u>	
	canB	(405)	----- <u>CAACAA</u> ACAGCC <u>AAAG</u> - <u>ATAGGGCC</u> <u>ACTG</u> ----- <u>CCCT</u>	
	CanC	(414)	----- <u>CACTAA</u> CAAG <u>ATA</u> <u>ACGCTA</u> - <u>ACGTTAGA</u> ----- <u>A</u> ----- <u>G</u> <u>CCCT</u>	
5	canD_partial	(326)	----- <u>TGGACAA</u> AT <u>GACGGC</u> <u>AA</u> CA <u>ATCA</u> <u>CC</u> <u>CAAGA</u> <u>TA</u> <u>AGGGTTG</u> <u>TAGCC</u>	
	canE_partial	(316)	----- <u>GACAA</u> <u>CC</u> <u>ACTGCC</u> <u>AGA</u> <u>TAGA</u> <u>CC</u> <u>CC</u> <u>ACC</u> ----- <u>CCCT</u>	
	Consensus	(501)	C ACAAC AG AAAG AGAAC A A T A	GCCT
			551	600
	CanA	(548)	<u>TCTAC</u> <u>CCAGG</u> <u>CTAACC</u> <u>AGG</u> <u>CTAT</u> <u>CC</u> <u>ATTC</u> <u>AC</u> <u>CC</u> <u>CT</u> <u>CC</u> <u>AG</u> <u>TG</u> <u>AT</u> <u>AT</u> <u>TC</u>	
	canB	(437)	<u>ACTAC</u> <u>CCG</u> <u>AC</u> <u>CTAACC</u> <u>AGG</u> <u>CC</u> <u>AT</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>CT</u> <u>AA</u> <u>AT</u> <u>AT</u> <u>TC</u>	
10	CanC	(446)	<u>ACTA</u> <u>TG</u> <u>AC</u> <u>CC</u> <u>CTAACC</u> <u>AGG</u> <u>CC</u> <u>AT</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>TA</u> <u>AT</u> <u>AC</u> <u>TC</u>	
	canD_partial	(371)	<u>ACTA</u> <u>TG</u> <u>AC</u> <u>CC</u> <u>CTAACC</u> <u>AGG</u> <u>CC</u> <u>AT</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>TA</u> <u>AT</u> <u>AC</u> <u>TC</u>	
	canE_partial	(347)	<u>ACTAC</u> <u>CCG</u> <u>CTAACC</u> <u>AGG</u> <u>CC</u> <u>AT</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>CT</u> <u>AC</u> <u>CC</u> <u>TA</u> <u>AT</u> <u>AC</u> <u>TC</u>	
	Consensus	(551)	ACTACGAGGCTAAGGAGGGTATGCTATTCGACAGCCT CC T ATA T	
			601	642
15	CanA	(598)	<u>AACT</u> <u>TCCAC</u> <u>GGCT</u> <u>ACAA</u> <u>STAGG</u> <u>CTAA</u> -----	
	canB	(487)	<u>AA</u> <u>CA</u> <u>TA</u> <u>AC</u> <u>CC</u> <u>GG</u> <u>CT</u> <u>AA</u> <u>AG</u> <u>CT</u> <u>CA</u> <u>G</u> <u>CT</u> <u>AA</u> -----	
	CanC	(496)	<u>AACT</u> <u>TCCAC</u> <u>GGCT</u> <u>AC</u> <u>AG</u> <u>CC</u> <u>GG</u> <u>CT</u> <u>AA</u> <u>CC</u> <u>AG</u> <u>TC</u> <u>CC</u> <u>CT</u> <u>GT</u> <u>GG</u> <u>TA</u>	
	canD_partial	(396)	-----	
	canE_partial	(373)	-----	
20	Consensus	(601)	AAC T CAGGT CT G T	

Amino Acid Alignment for SEQ ID NOS:2, 4, 6, 8, and 10:

1			50	
	CanA_pep	(1)	<u>V</u> <u>KY</u> <u>T</u> <u>LA</u> <u>TA</u> <u>M</u> <u>AG</u> <u>I</u> <u>Y</u> <u>AS</u> <u>AA</u> <u>AL</u> <u>AL</u> <u>LA</u> <u>GF</u> <u>AT</u> <u>T</u> <u>Q</u> <u>S</u> <u>P</u> <u>L</u> <u>N</u> <u>S</u> <u>F</u> <u>Y</u> <u>A</u> <u>T</u> <u>G</u> <u>T</u> <u>A</u> <u>Q</u> <u>A</u> <u>V</u> <u>S</u> <u>E</u> <u>P</u> <u>I</u> <u>D</u> <u>V</u> <u>E</u>	
25	CanB_pep	(1)	<u>V</u> <u>K</u> <u>P</u> <u>T</u> <u>LA</u> <u>TA</u> <u>M</u> <u>AG</u> <u>I</u> <u>Y</u> <u>AS</u> <u>AA</u> <u>DL</u> <u>AL</u> <u>AL</u> <u>LA</u> <u>GF</u> <u>AT</u> <u>T</u> <u>Q</u> <u>S</u> <u>P</u> <u>L</u> <u>N</u> <u>S</u> <u>F</u> <u>Y</u> <u>A</u> <u>T</u> <u>G</u> <u>T</u> <u>A</u> <u>A</u> <u>T</u> <u>S</u> <u>E</u> <u>P</u> <u>I</u> <u>D</u> <u>V</u> <u>E</u>	
	CanC_pep	(1)	<u>M</u> <u>RY</u> <u>T</u> <u>LA</u> <u>TA</u> <u>M</u> <u>AG</u> <u>I</u> <u>Y</u> <u>AS</u> <u>AA</u> <u>AL</u> <u>AL</u> <u>LA</u> <u>GF</u> <u>AT</u> <u>T</u> <u>Q</u> <u>S</u> <u>P</u> <u>L</u> <u>N</u> <u>S</u> <u>F</u> <u>Y</u> <u>A</u> <u>T</u> <u>G</u> <u>T</u> <u>A</u> <u>Q</u> <u>A</u> <u>V</u> <u>S</u> <u>E</u> <u>P</u> <u>I</u> <u>D</u> <u>V</u> <u>E</u>	
	CanD_partial	(1)	-----	<u>S</u> <u>F</u> <u>Y</u> <u>A</u> <u>T</u> <u>G</u> <u>T</u> <u>A</u> <u>Q</u> <u>A</u> <u>V</u> <u>S</u> <u>E</u> <u>P</u> <u>I</u> <u>D</u> <u>V</u> <u>E</u>
	CanE_partial	(1)	-----	<u>S</u> <u>F</u> <u>Y</u> <u>A</u> <u>T</u> <u>G</u> <u>T</u> <u>A</u> <u>E</u> <u>T</u> <u>S</u> <u>E</u> <u>P</u> <u>I</u> <u>D</u> <u>V</u> <u>E</u>
	Consensus	(1)	V K T L A L A G I I A S A A L A L L A G F A T T Q S P L S F Y A T G T A Q A V S E P I D V E	
30			51	100
	CanA_pep	(51)	<u>S</u> <u>H</u> <u>I</u> <u>G</u> <u>-</u> <u>S</u> <u>I</u> <u>T</u> <u>P</u> <u>A</u> <u>A</u> <u>G</u> <u>Q</u> <u>S</u> <u>D</u> <u>D</u> <u>I</u> <u>C</u> <u>Y</u> <u>A</u> <u>I</u> <u>W</u> <u>I</u> <u>K</u> <u>D</u> <u>Q</u> <u>V</u> <u>N</u> <u>V</u> <u>K</u> <u>V</u> <u>I</u> <u>L</u> <u>R</u> <u>N</u> <u>A</u> <u>E</u> <u>O</u> <u>L</u> <u>K</u> <u>P</u> <u>Y</u> <u>E</u>	
	CanB_pep	(51)	<u>S</u> <u>H</u> <u>I</u> <u>S</u> <u>-</u> <u>S</u> <u>I</u> <u>A</u> <u>P</u> <u>A</u> <u>G</u> <u>Q</u> <u>S</u> <u>Q</u> <u>D</u> <u>I</u> <u>C</u> <u>Y</u> <u>F</u> <u>N</u> <u>V</u> <u>T</u> <u>A</u> <u>K</u> <u>D</u> <u>Q</u> <u>V</u> <u>N</u> <u>V</u> <u>K</u> <u>V</u> <u>I</u> <u>L</u> <u>A</u> <u>N</u> <u>E</u> <u>O</u> <u>L</u> <u>K</u> <u>P</u> <u>Y</u> <u>E</u>	
	CanC_pep	(51)	<u>S</u> <u>H</u> <u>I</u> <u>D</u> <u>N</u> <u>-</u> <u>S</u> <u>I</u> <u>A</u> <u>P</u> <u>A</u> <u>G</u> <u>Q</u> <u>Y</u> <u>K</u> <u>D</u> <u>C</u> <u>Y</u> <u>I</u> <u>K</u> <u>T</u> <u>N</u> <u>Q</u> <u>S</u> <u>K</u> <u>V</u> <u>N</u> <u>V</u> <u>K</u> <u>V</u> <u>I</u> <u>L</u> <u>A</u> <u>N</u> <u>E</u> <u>O</u> <u>L</u> <u>K</u> <u>P</u> <u>Y</u> <u>E</u>	
	CanD_partial	(19)	<u>S</u> <u>S</u> <u>I</u> <u>G</u> <u>T</u> <u>L</u> <u>N</u> <u>T</u> <u>-</u> <u>A</u> <u>A</u> <u>G</u> <u>Q</u> <u>S</u> <u>K</u> <u>Q</u> <u>T</u> <u>L</u> <u>D</u> <u>I</u> <u>T</u> <u>I</u> <u>Y</u> <u>H</u> <u>N</u> <u>D</u> <u>V</u> <u>N</u> <u>T</u> <u>K</u> <u>V</u> <u>I</u> <u>L</u> <u>A</u> <u>N</u> <u>A</u> <u>O</u> <u>L</u> <u>R</u> <u>P</u> <u>Y</u> <u>E</u>	
35	CanE_partial	(19)	<u>C</u> <u>M</u> <u>I</u> <u>N</u> <u>T</u> <u>A</u> <u>I</u> <u>A</u> <u>P</u> <u>A</u> <u>G</u> <u>Q</u> <u>S</u> <u>V</u> <u>G</u> <u>I</u> <u>S</u> <u>T</u> <u>I</u> <u>E</u> <u>N</u> <u>K</u> <u>T</u> <u>D</u> <u>V</u> <u>N</u> <u>V</u> <u>K</u> <u>I</u> <u>L</u> <u>A</u> <u>N</u> <u>E</u> <u>O</u> <u>L</u> <u>K</u> <u>P</u> <u>Y</u> <u>E</u>	
	Consensus	(51)	SHL SIAPAAAGQGS DIGYI I K VNVVKLKVTLANAEQLKPYFK	
			101	150
	CanA_pep	(100)	<u>Y</u> <u>L</u> <u>Q</u> <u>I</u> <u>O</u> <u>T</u> <u>S</u> <u>G</u> <u>Y</u> <u>E</u> <u>T</u> <u>N</u> <u>S</u> <u>T</u> <u>A</u> <u>L</u> <u>G</u> <u>N</u> <u>F</u> <u>S</u> <u>E</u> <u>T</u> <u>K</u> <u>A</u> <u>V</u> <u>I</u> <u>S</u> <u>D</u> <u>E</u> <u>N</u> <u>S</u> <u>A</u> <u>V</u> <u>I</u> <u>V</u> <u>L</u> <u>D</u> <u>K</u> <u>E</u> <u>D</u> <u>I</u> <u>A</u> <u>V</u> <u>L</u> <u>P</u> <u>D</u> <u>K</u>	
	CanB_pep	(100)	<u>Y</u> <u>L</u> <u>Q</u> <u>I</u> <u>V</u> <u>L</u> <u>K</u> <u>S</u> <u>E</u> <u>V</u> <u>A</u> <u>D</u> ----- <u>E</u> <u>I</u> <u>K</u> <u>A</u> <u>V</u> <u>I</u> <u>S</u> <u>D</u> <u>E</u> <u>N</u> <u>S</u> <u>A</u> <u>V</u> <u>I</u> <u>V</u> <u>L</u> <u>D</u> <u>K</u> <u>E</u> <u>D</u> <u>I</u> <u>A</u> <u>V</u> <u>L</u> <u>P</u> <u>D</u> <u>K</u>	
40	CanC_pep	(101)	<u>Y</u> <u>L</u> <u>Q</u> <u>I</u> <u>V</u> <u>L</u> <u>T</u> <u>S</u> <u>N</u> <u>A</u> <u>T</u> <u>G</u> ----- <u>T</u> <u>D</u> <u>M</u> <u>Y</u> <u>K</u> <u>A</u> <u>V</u> <u>I</u> <u>S</u> <u>D</u> <u>E</u> <u>N</u> <u>S</u> <u>A</u> <u>V</u> <u>I</u> <u>V</u> <u>L</u> <u>D</u> <u>N</u> <u>D</u> <u>D</u> <u>S</u> <u>T</u> <u>N</u> ---	
	CanD_partial	(68)	<u>Y</u> <u>L</u> <u>Q</u> <u>I</u> <u>I</u> <u>K</u> <u>L</u> <u>V</u> <u>S</u> <u>L</u> <u>D</u> <u>S</u> <u>N</u> ----- <u>N</u> <u>E</u> <u>S</u> <u>E</u> <u>E</u> <u>X</u> <u>G</u> <u>M</u> <u>I</u> <u>H</u> <u>L</u> <u>W</u> <u>K</u> <u>Y</u> <u>A</u> <u>V</u> <u>I</u> <u>L</u> <u>D</u> <u>H</u> <u>E</u> <u>D</u> <u>F</u> <u>N</u> <u>N</u> <u>D</u> <u>I</u> <u>D</u> <u>---</u>	
	CanE_partial	(69)	<u>Y</u> <u>L</u> <u>Q</u> <u>I</u> <u>V</u> <u>L</u> <u>K</u> <u>S</u> <u>V</u> <u>D</u> <u>S</u> ----- <u>E</u> <u>I</u> <u>K</u> <u>A</u> <u>V</u> <u>I</u> <u>S</u> <u>L</u> <u>E</u> <u>K</u> <u>S</u> <u>A</u> <u>V</u> <u>I</u> <u>L</u> <u>D</u> <u>N</u> <u>E</u> <u>F</u> <u>Q</u> ---	
	Consensus	(101)	YLQIVL S S EIKAVISLDKPSAVIILD EDF	
			151	200
45	CanA_pep	(150)	TGYNTNTSIWVPGEPDKIIIVYNETKPVAL <u>I</u> <u>N</u> <u>F</u> <u>K</u> <u>A</u> <u>F</u> <u>Y</u> <u>P</u> <u>A</u> <u>K</u> <u>E</u> <u>G</u> <u>M</u> <u>L</u> <u>F</u> <u>D</u> <u>S</u> <u>L</u> <u>P</u> <u>V</u> <u>I</u> <u>F</u>	

CanB_pep	(139)	-----AKISATA Y MEAKEGMLFDSPV I
CanC_pep	(141)	-----KIQ----- I KVEAYMEAKEGMLFDSPVI
CanD_partial	(111)	-----N--DGNNDAKIRVV AY MEAKEGML -----
CanE_partial	(105)	-----GDNQQQIDAT AY MEAKEGML -----
5	Consensus	(151) A I AYYEAKEGMLFDSPV
		201 214
	CanA_pep	(200) NFQVLQVG-----
	CanB_pep	(163) NIQVLSVS-----
	CanC_pep	(166) NFQVLSAACSPW-----
10	CanD_partial	(132) -----
	CanE_partial	(125) -----
	Consensus	(201) N QVL

15 The polymer may have a shape of a short fiber, and therefore is also called
 “polymer fiber.” The polymer fiber is made from monomeric protein units, e.g. Can A:
 182 amino acids: MW = 19,830 Daltons, having a sequence of SED ID NO:2. The
 secondary structure of the protein may be mainly β -sheets.

20 The protein subunits in the polymer are arranged in a right-handed or left-
 handed, two-stranded helix. Occasionally, the polymer fibers made up of a three-handed
 helix may be observed. The periodicity (the distance of one helix turn to the next) of the
 polymer is 4.4 nm. The polymer has a unique quaternary structure. The polymer fiber
 has an outer diameter of 25 nm and inner diameter, 21 nm (in suspension). Under an
 electronic microscope, the dry negatively stained polymer fibers exhibit an outer diameter
 of 32 nm due to collapsing. Length of the polymer fiber is mostly between 3 and 5
 25 micrometers. Some of the polymer fibers may reach a length from 10 to 25 micrometers.
 The polymer fibers may form bundles of tens and hundreds of fibers with an overall
 diameter of 100 to 500 nm. Occasionally the bundle may reach an overall diameter of
 4,000 nm. The polymer fiber is at least stable up to 128°C.

30 CanA nanotubules of the invention can exhibit remarkable heat stability,
 e.g. temperatures up to about 150°C or 140°C. In one aspect, the nanotubules of the
 invention have heat stability in temperatures up to 128°C and stability in 2% SDS at
 100°C for at least 60 minutes. Purified recombinant CanB protein will also form
 nanotubular structures but they are less regular and not as heat stable as the nanotubules
 assembled from CanA. Purified, recombinant CanC does not self-assemble into
 35 nanotubules. Together, CanA (SEQ ID NO:2), CanB (SEQ ID NO:4), and CanC (SEQ
 ID NO:6) represent three very similar proteins that exhibit significantly different
 polymerization potentials *in vitro*, as summarized in Table 2:

Protein	CanA	CanB	CanC
CanA	100%		
CanB	60% Identical 64% Similar	100%	
CanC	55% Identical 62% Similar	68% Identical 77% Similar	100%

Table 2: Comparison of amino acid sequences of CanA, CanB, CanC.

One difference between CanA and CanB is the 14 amino acid insertion near the middle of the CanA sequence (see Figure 2). Immunoelectron microscopy and an antibody specific for this 14 amino acid sequence have been used to determine that this 5 sequence is displayed on the surface of the assembled nanotubule. The absence of this corresponding sequence in CanB demonstrates that this peptide domain is nonessential for nanotubule assembly. Therefore, it is possible to remove this sequence and replace it with a peptide domain that alters the structure of CanA. In one aspect, replacing the endogenous 14 residue motif with a heterologous peptide changes the enantioselectivity 10 of CanA.

10 Recombinant chimeric proteins of the invention can be expressed in a cell, e.g., a bacteria, such as *E. coli*, and purified away from host proteins by using heat treatment to denature and precipitate (e.g., *E. coli*) protein. The soluble heat stable protein (e.g., CanA) can be recovered from the supernatant following centrifugation. The 15 chimeric protein can be assembly-competent at this stage. In one aspect, the self-assembly reaction is initiated by addition of millimolar concentrations of Ca^{++} and Mg^{++} . In one aspect, following assembly of the nanotubules, they are stable in cation-free buffer and buffers containing up to 20mM chelator, e.g., EDTA, EGTA.

20 *Colloidal Stability.* Nanotubules of the invention can interact at different levels by pairing, bundling, entangling (excluded volume interaction) and electrostatic cross-linking (bridging by divalent cations). The different types of aggregates have an increasing dimensionality from a pair of rods to an interconnected network. The bundling of CanA nanotubules appears to be a magnesium-dependent process. In the absence of magnesium, CanA displays minimal bundling. However, upon the addition of millimolar 25 concentrations of magnesium, CanA nanotubules will form bundles visible by standard phase contrast light microscopy.

Nanotubule Stiffness. CanA nanotubules have been imaged under the transmission electron (TEM) and atomic force microscopes (AFM). From analyses of thermal vibrations of a single fiber in vacuum under the TEM, it was found that the CanA bending modulus is about 5 ± 2 GPa. This result is somewhat greater than other rigid 5 biopolymers of the same dimensions, such as microtubules which have a bending modulus of nearly 2 GPa, and comparable to the bending moduli of the strongest synthetic polymer fibers like Poly(6-amide) or Poly(methylmethacrylate), or, PLEXIGLASTTM.

In one aspect, the chimeric CanA proteins of the invention are used as 10 chiral selectors, e.g., in capillary electrophoresis. Serum albumin was one of the first proteins used as a chiral stationary phase for the successful separation of enantiomers, see, e.g., (Allenmark, 1998). Numerous proteins have been used to accomplish many enantioseparations using capillary electrophoresis methods. These proteins include α_1 -acid glycoprotein, avidin, ovomucoid, transferrin, cytochrome c, lysozyme, pepsin, 15 cellulase, and cellobiohydrolase see, e.g., Tanada (2001) *supra*. Proteins are favorable for use as chiral selectors because they frequently can be used for a wide variety of enantioseparations, see, e.g., Lloyd (1995) *J. of Chromatography A* 694:285-296. In addition, because proteins can be used for chiral separations in aqueous buffers, they are a 20 good choice for the analysis of samples derived from biological material, see, e.g., Busch (1993) *supra*. Accordingly, in alternative aspects, the chimeric CanA polypeptides of the invention comprise chiral selection motifs from serum albumin, α_1 -acid glycoprotein, avidin, ovomucoid, transferrin, cytochrome c, lysozyme, pepsin, cellulase and cellobiohydrolase. The chimeric CanA of the invention can comprise any peptide motif 25 having a chiral selection capability. These motifs can be inserted into a CanA or added to a CanA. In one aspect, they are used to replace a subsequence of CanA that has been removed, e.g., a 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 or a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2.

A chimeric monomer or polymer of the invention can comprise a 30 detectable moiety. In one aspect, the heterologous motif is a detectable moiety, e.g., a green fluorescent protein. In one aspect, the invention provides a nanotubule comprising chimeric monomers comprising green fluorescent protein motifs. These monomers and nanotubules can be used to study nanotubule formation, dissolution and function. For example, Figure 3 is an illustration of an immunofluorescent light microscope image of

nanotubules assembled from a fusion protein generated by fusing the CanA open reading frame (SEQ ID NO:1) to the open reading frame of the green fluorescent protein ZSGREEN™ (BD Biosciences Clontech, Palo Alto, CA).

The invention provides enantioseparation methods using proteins free in solution as buffer additives, as described, e.g., in Busch (1993) *supra*, and using proteins immobilized by a variety of methods, as described, e.g., in Tanaka (2001) *supra*; Ito (2001) *J. of Chromatography A* 925:41-47. There are advantages and disadvantages to both approaches. By using proteins in solution, the native conformation of the protein is maintained resulting in a more uniform presentation of the sites involved in generating chiral resolution. However, in capillary electrophoresis-based methods, the presence of protein in the buffer solutions can produce extremely high background UV absorption. This limitation has been addressed by using partial filling and countercurrent techniques that allow relatively high concentration protein solutions to be used without causing background problems at the detector. Partial filling and countercurrent techniques are well known in the art, as, e.g., described in Tanaka (2001) *supra*; Chankvetadze (2001) *supra*.

In contrast, the use of immobilization techniques allows for the production of capillaries with high concentrations of the protein chiral selector. The potential drawback to these approaches is the heterogeneity introduced by the method of protein immobilization. This heterogeneity is particularly important when analyzing protein-ligand interactions (see, e.g., Lloyd (1995) *supra*). Changes in protein conformation introduced as a result of the immobilization method can significantly alter protein-ligand interactions and these types of analyses are therefore more often performed using protein chiral selectors in free solution.

Given their capacity for stereospecific molecular recognition (see, e.g., Lakshmi (1997) *Nature* 388:758-760; Henriksson (1996) *supra*), enzymes and apoenzymes are a source of chiral selectors used in the compositions and methods of the invention. Thus, the invention provides chimeric monomers and polymers, including nanotubules, comprising chiral selector enzymes and apoenzymes and chiral selector peptide motifs of enzymes and apoenzymes, such as enzyme active site motifs. The chimeric monomers and polymers, including nanotubules, of the invention can comprise any enzymes or apoenzymes, or any enzyme active site motif. For example, the chimeric monomers and polymers, including nanotubules, and active site motifs of the invention can be derived from glycosyltransferases, glycosylhydrolases, nitrilases, esterases,

amidases, lipases, polymerases, cellulases, hydrolases, deaminases, nitroreductases and the like.

Polypeptides and peptide for making and/or using the chimeric monomers and polymers of the invention can be isolated from natural sources, be synthetic, or be recombinantly generated polypeptides. Peptides and proteins can be recombinantly expressed *in vitro* or *in vivo*. The peptides and polypeptides of the invention can be made and isolated using any method known in the art. Polypeptide and peptides for making and/or using the chimeric monomers and polymers of the invention can also be synthesized, whole or in part, using chemical methods well known in the art. See e.g., Caruthers (1980) Nucleic Acids Res. Symp. Ser. 215-223; Horn (1980) Nucleic Acids Res. Symp. Ser. 225-232; Banga, A.K., Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems (1995) Technomic Publishing Co., Lancaster, PA. For example, peptide synthesis can be performed using various solid-phase techniques (see e.g., Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) and automated synthesis may be achieved, e.g., using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The peptides and polypeptides for making and/or using the chimeric monomers and polymers of the invention can also be glycosylated. The glycosylation can be added post-translationally either chemically or by cellular biosynthetic mechanisms, wherein the later incorporates the use of known glycosylation motifs, which can be native to the sequence or can be added as a peptide or added in the nucleic acid coding sequence. The glycosylation can be O-linked or N-linked.

The peptides and polypeptides for making and/or using the chimeric monomers and polymers of the invention, as defined above, include all "mimetic" and "peptidomimetic" forms. The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound which has substantially the same structural and/or functional characteristics of the polypeptides of the invention. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. As with polypeptides of the invention which are conservative variants, routine experimentation will determine whether a mimetic is within the scope of the invention,

i.e., that its structure and/or function is not substantially altered. Thus, in one aspect, a mimetic composition is within the scope of the invention if it has an amylase activity.

Polypeptide mimetic compositions can contain any combination of non-natural structural components. In alternative aspect, mimetic compositions of the invention include one or all of the following three structural groups: a) residue linkage groups other than the natural amide bond ("peptide bond") linkages; b) non-natural residues in place of naturally occurring amino acid residues; or c) residues which induce secondary structural mimicry, i.e., to induce or stabilize a secondary structure, e.g., a beta turn, gamma turn, beta sheet, alpha helix conformation, and the like. For example, a polypeptide of the invention can be characterized as a mimetic when all or some of its residues are joined by chemical means other than natural peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as, e.g., glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-diisopropylcarbodiimide (DIC). Linking groups that can be an alternative to the traditional amide bond ("peptide bond") linkages include, e.g., ketomethylene (e.g., -C(=O)-CH₂- for -C(=O)-NH-), aminomethylene (CH₂-NH), ethylene, olefin (CH=CH), ether (CH₂-O), thioether (CH₂-S), tetrazole (CN₄-), thiazole, retroamide, thioamide, or ester (see, e.g., Spatola (1983) in Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY).

A polypeptide can also be characterized as a mimetic by containing all or some non-natural residues in place of naturally occurring amino acid residues. Non-natural residues are well described in the scientific and patent literature; a few exemplary non-natural compositions useful as mimetics of natural amino acid residues and guidelines are described below. Mimetics of aromatic amino acids can be generated by replacing by, e.g., D- or L- naphylalanine; D- or L- phenylglycine; D- or L-2 thieneylalanine; D- or L-1, -2, 3-, or 4- pyreneylalanine; D- or L-3 thieneylalanine; D- or L-(2-pyridinyl)-alanine; D- or L-(3-pyridinyl)-alanine; D- or L-(2-pyrazinyl)-alanine; D- or L-(4-isopropyl)-phenylglycine; D-(trifluoromethyl)-phenylglycine; D-(trifluoromethyl)-phenylalanine; D-p-fluoro-phenylalanine; D- or L-p-biphenylphenylalanine; D- or L-p-methoxy-biphenylphenylalanine; D- or L-2-indole(alkyl)alanines; and, D- or L-alkylalanines, where alkyl can be substituted or unsubstituted methyl, ethyl, propyl, hexyl, butyl, pentyl, isopropyl, iso-butyl, sec-isotyl, iso-pentyl, or a non-acidic amino acids. Aromatic rings of a non-natural amino acid

include, e.g., thiazolyl, thiophenyl, pyrazolyl, benzimidazolyl, naphthyl, furanyl, pyrrolyl, and pyridyl aromatic rings.

Mimetics of acidic amino acids can be generated by substitution by, e.g., non-carboxylate amino acids while maintaining a negative charge; (phosphono)alanine; sulfated threonine. Carboxyl side groups (e.g., aspartyl or glutamyl) can also be selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as, e.g., 1-cyclohexyl-3(2-morpholiny1-(4-ethyl) carbodiimide or 1-ethyl-3(4-azonia- 4,4-dimethylpentyl) carbodiimide. Aspartyl or glutamyl can also be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions. Mimetics of basic amino acids can be generated by substitution with, e.g., (in addition to lysine and arginine) the amino acids ornithine, citrulline, or (guanidino)-acetic acid, or (guanidino)alkyl-acetic acid, where alkyl is defined above. Nitrile derivative (e.g., containing the CN-moiety in place of COOH) can be substituted for asparagine or glutamine. Asparaginyl and glutaminyl residues can be deaminated to the corresponding aspartyl or glutamyl residues. Arginine residue mimetics can be generated by reacting arginyl with, e.g., one or more conventional reagents, including, e.g., phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, or ninhydrin, preferably under alkaline conditions. Tyrosine residue mimetics can be generated by reacting tyrosyl with, e.g., aromatic diazonium compounds or tetranitromethane. N-acetylimidizol and tetranitromethane can be used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively. Cysteine residue mimetics can be generated by reacting cysteinyl residues with, e.g., alpha-haloacetates such as 2-chloroacetic acid or chloroacetamide and corresponding amines; to give carboxymethyl or carboxyamidomethyl derivatives. Cysteine residue mimetics can also be generated by reacting cysteinyl residues with, e.g., bromo-trifluoroacetone, alpha-bromo-beta-(5-imidozoyl) propionic acid; chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide; methyl 2-pyridyl disulfide; p-chloromercuribenzoate; 2-chloromercuri-4-nitrophenol; or, chloro-7-nitrobenzo-oxa-1,3-diazole. Lysine mimetics can be generated (and amino terminal residues can be altered) by reacting lysinyl with, e.g., succinic or other carboxylic acid anhydrides. Lysine and other alpha-amino-containing residue mimetics can also be generated by reaction with imidoesters, such as methyl picolinimidate, pyridoxal phosphate, pyridoxal, chloroborohydride, trinitrobenzenesulfonic acid, O-methylisourea, 2,4, pentanedione, and transamidase-catalyzed reactions with glyoxylate. Mimetics of methionine can be generated by reaction with, e.g., methionine sulfoxide. Mimetics of proline include, e.g., pipecolic acid, thiazolidine

carboxylic acid, 3- or 4- hydroxy proline, dehydroproline, 3- or 4-methylproline, or 3,3,-dimethylproline. Histidine residue mimetics can be generated by reacting histidyl with, e.g., diethylprocarbонate or para-bromophenacyl bromide. Other mimetics include, e.g., those generated by hydroxylation of proline and lysine; phosphorylation of the hydroxyl groups of seryl or threonyl residues; methylation of the alpha-amino groups of lysine, arginine and histidine; acetylation of the N-terminal amine; methylation of main chain amide residues or substitution with N-methyl amino acids; or amidation of C-terminal carboxyl groups.

A residue, e.g., an amino acid, of a polypeptide for making and/or using the chimeric monomers and polymers of the invention can also be replaced by an amino acid (or peptidomimetic residue) of the opposite chirality. Thus, any amino acid naturally occurring in the L-configuration (which can also be referred to as the R or S, depending upon the structure of the chemical entity) can be replaced with the amino acid of the same chemical structural type or a peptidomimetic, but of the opposite chirality, referred to as the D- amino acid, but also can be referred to as the R- or S- form.

The invention also provides methods for modifying the chimeric polypeptides of the invention by either natural processes, such as post-translational processing (e.g., phosphorylation, acylation, etc), or by chemical modification techniques, and the resulting modified polypeptides. Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide may have many types of modifications. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. See, e.g., Creighton, T.E., Proteins – Structure and Molecular Properties 2nd Ed., W.H. Freeman and

Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983).

Solid-phase chemical peptide synthesis methods can also be used to synthesize the polypeptide or fragments for making and/or using the chimeric monomers and polymers of the invention. Such method are known in the art, see, e.g., Merrifield 5 (1963) J. Am. Chem. Soc. 85:2149-2154; Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12; and have been employed in commercially available laboratory peptide design and synthesis kits (Cambridge Research Biochemicals). Such commercially available laboratory kits have 10 generally utilized the teachings of H. M. Geysen et al, Proc. Natl. Acad. Sci., USA, 81:3998 (1984) and provide for synthesizing peptides upon the tips of a multitude of "rods" or "pins" all of which are connected to a single plate. When such a system is utilized, a plate of rods or pins is inverted and inserted into a second plate of corresponding wells or reservoirs, which contain solutions for attaching or anchoring an 15 appropriate amino acid to the pin's or rod's tips. By repeating such a process step, i.e., inverting and inserting the rod's and pin's tips into appropriate solutions, amino acids are built into desired peptides. In addition, a number of available Fmoc peptide synthesis systems are available.

Generating and Manipulating Nucleic Acids

20 The invention provides nucleic acids, including expression cassettes such as expression vectors, encoding the chimeric polypeptides of the invention. The invention also includes methods for modifying nucleic acids encoding the chimeric polypeptides of the invention by, e.g., synthetic ligation reassembly, optimized directed evolution system and/or saturation mutagenesis.

25 The nucleic acids of the invention can be made, isolated and/or manipulated by, e.g., cloning and expression of cDNA libraries, amplification of message or genomic DNA by PCR, and the like. The invention can be practiced in conjunction with any method or protocol or device known in the art, which are well described in the scientific and patent literature.

30 General Techniques

The nucleic acids used to practice this invention, whether RNA, tRNA, antisense nucleic acid, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed/

generated recombinantly. Recombinant polypeptides generated from these nucleic acids can be individually isolated or cloned and tested for a desired activity. Any recombinant expression system can be used, including bacterial, mammalian, yeast, insect or plant cell expression systems.

5 Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Patent No. 4,458,066.

10 Techniques for the manipulation of nucleic acids, such as, e.g., subcloning, labeling probes (e.g., random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, hybridization and the like are well described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A
15 LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. 20 Elsevier, N.Y. (1993).

25 Another useful means of obtaining and manipulating nucleic acids used to practice the invention is to clone from genomic samples, and, if desired, screen and re-clone inserts isolated or amplified from, e.g., genomic clones or cDNA clones. Sources of nucleic acid used in practicing the invention include genomic or cDNA libraries contained in, e.g., mammalian artificial chromosomes (MACs), see, e.g., U.S. Patent Nos. 5,721,118; 6,025,155; human artificial chromosomes, see, e.g., Rosenfeld (1997) Nat. Genet. 15:333-335; yeast artificial chromosomes (YAC); bacterial artificial chromosomes (BAC); P1 artificial chromosomes, see, e.g., Woon (1998) Genomics 50:306-316; P1-derived vectors (PACs), see, e.g., Kern (1997) Biotechniques 23:120-124; cosmids, recombinant viruses, phages or plasmids.

30 In one aspect, a nucleic acid encoding a polypeptide of the invention is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptide or fragment thereof.

The invention provides fusion proteins and nucleic acids encoding them. In addition to chiral selection motifs, enzymes, receptors, ligands, antibodies, antigens, epitopes and the like, polypeptide of the invention can be fused to a heterologous peptide or polypeptide such as N-terminal identification peptide, which imparts desired characteristics such as increased stability or simplified purification. Peptides and polypeptides of the invention also can be synthesized and expressed as fusion proteins with one or more additional domains linked thereto for, e.g., producing a more immunogenic peptide, to more readily isolate a recombinantly synthesized peptide, to identify and isolate antibodies and antibody-expressing B cells, and the like. Detection and purification facilitating domains include, e.g., metal chelating peptides such as polyhistidine tracts and histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego CA) between a purification domain and the motif-comprising peptide or polypeptide to facilitate purification. For example, an expression vector can include an epitope-encoding nucleic acid sequence linked to six histidine residues followed by a thioredoxin and an enterokinase cleavage site (see e.g., Williams (1995) Biochemistry 34:1787-1797; Dobeli (1998) Protein Expr. Purif. 12:404-414). The histidine residues facilitate detection and purification while the enterokinase cleavage site provides a means for purifying the epitope from the remainder of the fusion protein.

Transcriptional and translational control sequences

The invention provides nucleic acid (e.g., DNA) sequences of the invention operatively linked to expression (e.g., transcriptional or translational) control sequence(s), e.g., promoters or enhancers, to direct or modulate RNA synthesis/ expression. The expression control sequence can be in an expression vector. Exemplary bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Exemplary eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein I.

Promoters suitable for expressing a polypeptide in bacteria include the *E. coli* lac or trp promoters, the lacI promoter, the lacZ promoter, the T3 promoter, the T7 promoter, the gpt promoter, the lambda PR promoter, the lambda PL promoter, promoters

from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses, and the mouse metallothionein-I promoter.

5 Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used.

The invention provides expression cassettes that can be expressed in a tissue-specific manner, e.g., that can express a chimeric polypeptide of the invention in a tissue-specific manner. The invention provides plants or seeds that express a chimeric 10 polypeptide of the invention in a tissue-specific manner. The tissue-specificity can be seed specific, stem specific, leaf specific, root specific, fruit specific and the like. The nucleic acids of the invention can also be operably linked to plant promoters which are inducible upon exposure to chemicals reagents.

Expression vectors and cloning vehicles

15 The invention provides expression vectors and cloning vehicles comprising nucleic acids of the invention, e.g., sequences encoding the chimeric polypeptides of the invention. Expression vectors and cloning vehicles of the invention can comprise viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral DNA (e.g., vaccinia, adenovirus, foul pox virus, 20 pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus, Aspergillus and yeast). Vectors of the invention can include chromosomal, non-chromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and are commercially available.

25 Exemplary vectors are include: bacterial: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, (lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as they are replicable and viable in the host. Low copy number or high copy number vectors may be 30 employed with the present invention.

The expression vector can comprise a promoter, a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Mammalian expression vectors can

comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences. In some aspects, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required non-transcribed genetic elements.

5 In one aspect, the expression vectors contain one or more selectable marker genes to permit selection of host cells containing the vector. Such selectable markers include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli*, and the *S. cerevisiae* TRP1 gene. Promoter regions can be selected from any desired gene 10 using chloramphenicol transferase (CAT) vectors or other vectors with selectable markers.

15 Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells can also contain enhancers to increase expression levels. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and the adenovirus enhancers.

20 A nucleic acid sequence can be inserted into a vector by a variety of procedures. In general, the sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, blunt ends in both the insert and the vector may be ligated. A variety of cloning techniques are known in the art, e.g., as described in Ausubel and 25 Sambrook. Such procedures and others are deemed to be within the scope of those skilled in the art.

25 The vector can be in the form of a plasmid, a viral particle, or a phage. Other vectors include chromosomal, non-chromosomal and synthetic DNA sequences, derivatives of SV40; bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. A variety of cloning and expression 30 vectors for use with prokaryotic and eukaryotic hosts are described by, e.g., Sambrook. Any vector may be used as long as it is replicable and viable in the host cell.

The nucleic acids of the invention can be expressed in expression cassettes, vectors or viruses and transiently or stably expressed in plant cells and seeds. One exemplary transient expression system uses episomal expression systems, e.g.,

cauliflower mosaic virus (CaMV) viral RNA generated in the nucleus by transcription of an episomal mini-chromosome containing supercoiled DNA, see, e.g., Covey (1990)

Proc. Natl. Acad. Sci. USA 87:1633-1637. Alternatively, coding sequences, i.e., all or sub-fragments of sequences of the invention can be inserted into a plant host cell genome

5 becoming an integral part of the host chromosomal DNA. Sense or antisense transcripts can be expressed in this manner. A vector comprising the sequences (e.g., promoters or coding regions) from nucleic acids of the invention can comprise a marker gene that confers a selectable phenotype on a plant cell or a seed. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to

10 kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

Expression vectors capable of expressing nucleic acids and proteins in plants are well known in the art, and can include, e.g., vectors from *Agrobacterium* spp., potato virus X (see, e.g., Angell (1997) EMBO J. 16:3675-3684), tobacco mosaic virus

15 (see, e.g., Casper (1996) Gene 173:69-73), tomato bushy stunt virus (see, e.g., Hillman (1989) Virology 169:42-50), tobacco etch virus (see, e.g., Dolja (1997) Virology 234:243-252), bean golden mosaic virus (see, e.g., Morinaga (1993) Microbiol Immunol. 37:471-476), cauliflower mosaic virus (see, e.g., Cecchini (1997) Mol. Plant Microbe

Interact. 10:1094-1101), maize Ac/Ds transposable element (see, e.g., Rubin (1997) Mol. 20 Cell. Biol. 17:6294-6302; Kunze (1996) Curr. Top. Microbiol. Immunol. 204:161-194), and the maize suppressor-mutator (Spm) transposable element (see, e.g., Schlappi (1996) Plant Mol. Biol. 32:717-725); and derivatives thereof.

In one aspect, the expression vector can have two replication systems to allow it to be maintained in two organisms, for example in mammalian or insect cells for

25 expression and in a prokaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector can contain at least one sequence homologous to the host cell genome. It can contain two homologous sequences which flank the expression construct. The integrating vector can be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the

30 vector. Constructs for integrating vectors are well known in the art.

Expression vectors of the invention may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed, e.g., genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers can also

include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

The terms "vector" and "expression cassette" as used herein can be used interchangeably and refer to a nucleotide sequence which is capable of affecting expression of a nucleic acid, e.g., a mutated nucleic acid of the invention. Expression cassettes can include at least a promoter operably linked with the polypeptide coding sequence; and, optionally, with other sequences, e.g., transcription termination signals. Additional factors necessary or helpful in effecting expression may also be used, e.g., enhancers. "Operably linked" as used herein refers to linkage of a promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence. Thus, expression cassettes also include plasmids, expression vectors, recombinant viruses, any form of recombinant "naked DNA" vector, and the like. A "vector" comprises a nucleic acid which can infect, transfect, transiently or permanently transduce a cell. It will be recognized that a vector can be a naked nucleic acid, or a nucleic acid complexed with protein or lipid. The vector optionally comprises viral or bacterial nucleic acids and/or proteins, and/or membranes (e.g., a cell membrane, a viral lipid envelope, etc.). Vectors include, but are not limited to replicons (e.g., RNA replicons, bacteriophages) to which fragments of DNA may be attached and become replicated. Vectors thus include, but are not limited to RNA, autonomous self-replicating circular or linear DNA or RNA (e.g., plasmids, viruses, and the like, see, e.g., U.S. Patent No. 5,217,879), and includes both the expression and non-expression plasmids.

Host cells and transformed cells

The invention also provides a transformed cell comprising a nucleic acid sequence of the invention, e.g., a sequence encoding chimeric polypeptides of the invention, or an expression cassette, e.g., a vector, of the invention. The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells, eukaryotic cells, such as bacterial cells, fungal cells, yeast cells, mammalian cells, insect cells, or plant cells. Exemplary bacterial cells include *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Exemplary insect cells include *Drosophila S2* and *Spodoptera Sf9*. Exemplary animal cells include CHO, COS or Bowes melanoma or any mouse or human cell line. The selection of an appropriate host is within the abilities of those skilled in the art. Techniques for transforming a wide variety of higher plant

species are well known and described in the technical and scientific literature. See, e.g., Weising (1988) *Ann. Rev. Genet.* 22:421-477, U.S. Patent No. 5,750,870.

The vector can be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, J., *Basic Methods in Molecular Biology*, (1986)).

In one aspect, the nucleic acids or vectors of the invention are introduced into the cells for screening, thus, the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type. Exemplary methods include CaPO₄ precipitation, liposome fusion, lipofection (e.g., LIPOFECTIN™), electroporation, viral infection, etc. The candidate nucleic acids may stably integrate into the genome of the host cell (for example, with retroviral introduction) or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.). As many pharmaceutically important screens require human or model mammalian cell targets, retroviral vectors capable of transfecting such targets are preferred.

Cells can be harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Depending upon the host employed in a recombinant production procedure, the polypeptides produced by host cells containing the vector may be glycosylated or may be non-glycosylated.

5 Polypeptides of the invention may or may not also include an initial methionine amino acid residue.

Cell-free translation systems can also be employed to produce a polypeptide of the invention. Cell-free translation systems can use mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment thereof. In some aspects, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

10 The expression vectors can contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Amplification of Nucleic Acids

15 In practicing the invention, nucleic acids of the invention and nucleic acids encoding the chimeric polypeptides of the invention, or modified nucleic acids of the invention, can be reproduced by amplification. Amplification can also be used to clone or modify the nucleic acids of the invention. Thus, the invention provides amplification primer sequence pairs for amplifying nucleic acids of the invention.

20 Amplification reactions can also be used to quantify the amount of nucleic acid in a sample (such as the amount of message in a cell sample), label the nucleic acid (e.g., to apply it to an array or a blot), detect the nucleic acid, or quantify the amount of a specific nucleic acid in a sample. In one aspect of the invention, message isolated from a cell or a cDNA library are amplified.

25 The skilled artisan can select and design suitable oligonucleotide amplification primers. Amplification methods are also well known in the art, and include, e.g., polymerase chain reaction, PCR (see, e.g., PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR)

(see, e.g., Wu (1989) *Genomics* 4:560; Landegren (1988) *Science* 241:1077; Barringer (1990) *Gene* 89:117); transcription amplification (see, e.g., Kwok (1989) *Proc. Natl. Acad. Sci. USA* 86:1173); and, self-sustained sequence replication (see, e.g., Guatelli (1990) *Proc. Natl. Acad. Sci. USA* 87:1874); Q Beta replicase amplification (see, e.g., 5 Smith (1997) *J. Clin. Microbiol.* 35:1477-1491), automated Q-beta replicase amplification assay (see, e.g., Burg (1996) *Mol. Cell. Probes* 10:257-271) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario); see also Berger (1987) *Methods Enzymol.* 152:307-316; Sambrook; Ausubel; U.S. Patent Nos. 4,683,195 and 4,683,202; Sooknanan (1995) *Biotechnology* 13:563-564.

10 Determining the degree of sequence identity

The cannulae polypeptide can comprise a protein having at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or more, sequence identity to SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, and is capable of assembling into a polymer, e.g., a nanotubule, or, is capable of acting as a chiral selector. The 15 chimeric cannulae proteins can assemble into nanotubular polymers to act as a chiral selectors, biosynthetic pathways, selection scaffoldings and the like. The extent of sequence identity (homology) may be determined using any computer program and associated parameters, including those described herein, such as BLAST 2.2.2. or FASTA version 3.0t78, with the default parameters.

20 Various sequence comparison programs identified herein are used in this aspect of the invention. Protein and/or nucleic acid sequence identities (homologies) may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are not limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448, 1988; Altschul et al., *J. Mol. Biol.* 215(3):403-410, 1990; Thompson et al., *Nucleic Acids Res.* 22(2):4673-4680, 1994; Higgins et al., 25 *Methods Enzymol.* 266:383-402, 1996; Altschul et al., *J. Mol. Biol.* 215(3):403-410, 1990; Altschul et al., *Nature Genetics* 3:266-272, 1993).

30 Homology or identity can be measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various deletions, substitutions and other modifications. The terms "homology" and "identity" in

the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region as measured using any

5 number of sequence comparison algorithms or by manual alignment and visual inspection. For sequence comparison, one sequence can act as a reference sequence (a sequence of the invention to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program
10 parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

15 A “comparison window”, as used herein, includes reference to a segment of any one of the numbers of contiguous residues. For example, in alternative aspects of the invention, contiguous residues ranging anywhere from 20 to the full length of an exemplary polypeptide or nucleic acid sequence of the invention are compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. If the reference sequence has the requisite sequence identity to an
20 exemplary polypeptide or nucleic acid sequence of the invention, e.g., 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90% or 95%, 98%, 99% or more sequence identity to a cannulae polypeptide, that sequence may be within the scope of the invention. In alternative embodiments, subsequences ranging from about 20 to 600, about 50 to 200, and about 100 to 150 are compared to a reference sequence of the same number of contiguous
25 positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981, by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970, by the search for similarity method of
30 person & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, in addition to a BLAST program

(Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (BLOCKs IMPROved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome databases to identify 5 polynucleotide sequences having substantially identical sequences. A number of genome databases are available, for example, a substantial portion of the human genome is 10 available as part of the Human Genome Sequencing Project (Gibbs, 1995). Databases containing genomic information annotated with some functional information are 15 maintained by different organization, and are accessible via the internet.

BLAST, BLAST 2.0 and BLAST 2.2.2 algorithms are also used to practice 20 the invention. They are described, e.g., in Altschul (1977) Nuc. Acids Res. 25:3389-3402; Altschul (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by 25 identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul (1990) *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along 30 each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its

maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) 5 uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N= -4, and a comparison of both 10 strands. The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match 15 between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001. In one aspect, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST"). For example, five specific BLAST programs can be 20 used to perform the following task: (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database; (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database; (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database; (4) TBLASTN compares a query protein sequence 25 against a nucleotide sequence database translated in all six reading frames (both strands); and, (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid 30 sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., Science 256:1443-1445, 1992; Henikoff and Henikoff, Proteins 17:49-61, 1993). Less preferably, the PAM or PAM250

matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation).

In one aspect of the invention, to determine if a nucleic acid has the requisite sequence identity to be within the scope of the invention, the NCBI BLAST 2.2.2 programs is used, default options to blastp. There are about 38 setting options in the BLAST 2.2.2 program. In this exemplary aspect of the invention, all default values are used except for the default filtering setting (i.e., all parameters set to default except filtering which is set to OFF); in its place a "-F F" setting is used, which disables filtering.

10 Use of default filtering often results in Karlin-Altschul violations due to short length of sequence.

The default values used in this exemplary aspect of the invention include:

"Filter for low complexity: ON
Word Size: 3
15 Matrix: Blosum62
Gap Costs: Existence:11
Extension:1"

Other default settings can be: filter for low complexity OFF, word size of 3 for protein, BLOSUM62 matrix, gap existence penalty of -11 and a gap extension penalty of -1. An exemplary NCBI BLAST 2.2.2 program setting has the "-W" option default to 0. This means that, if not set, the word size defaults to 3 for proteins and 11 for nucleotides.

Modification of Nucleic Acids

The invention provides methods of generating variants of the nucleic acids encoding the chimeric polypeptides of the invention. These methods can be repeated or used in various combinations to generate chimeric polypeptides having an altered or different activity or an altered or different stability from that of a chimeric polypeptide encoded by the template nucleic acid. These methods also can be repeated or used in various combinations, e.g., to generate variations in gene/ message expression, message translation or message stability. In another aspect, the genetic composition of a cell is altered by, e.g., modification of a homologous gene *ex vivo*, followed by its reinsertion into the cell.

The invention provides methods for evolving enzymes *in vitro* or *in vivo* to produce variants with characteristics tailored for specific applications. For example, using the evolution strategies of the invention, enzyme active sites can be modified to produce proteins that retain stereospecific substrate recognition but lack catalytic activity.

5 In one aspect, the chimeric monomers and polymers of the invention are evolved for applications in chiral selection using targeted mutagenesis and *in vitro* evolution strategies, e.g., as described herein, such as Gene Site Saturation Mutagenesis (GSSM™) and GeneReassembly™ (see, e.g., U.S. Patents 6,171,820, and 5,965,408 respectively). These technologies are used to create large libraries of mutagenized sequence variants
10 that are screened in a high throughput (HT) assay that selects mutants with a desired phenotype.

With GSSM™, the effects of all 64 codons (even nonsense codons) can be tested at each triplet position along the entire length of the open reading frame of the gene being analyzed. For example, in the case of a 200 amino acid protein, the gene can be
15 simultaneously assembled in 200 different reaction tubes where all 64 codons are present during the synthesis of each amino acid. The result is a library of single point mutants with all possible codons represented at each position of the open reading frame. The library of GSSM™ variants then can be screened using a HT assay to identify variants that have evolved the target phenotype. Individual GSSM™ variants that exhibit the
20 desired property then can be further evolved using GeneReassembly™.

In GeneReassembly™, a new library of mutants can be constructed by recombining DNA fragments taken from the single point mutant sequences identified in the GSSM screen. Therefore, the reassembly library can contain open reading frames that contain multiple point mutations that have accumulated as a result of the recombination process. The reassembled variants can be screened to identify mutant combinations with further improvements in the target activity. If necessary, GeneReassembly™ can be repeated until an evolved protein with the desired target properties is identified. These protein evolution strategies do not require prior knowledge of protein structure and therefore produce unbiased pools of protein variants for screening.
25

30 In one aspect, the invention provides combinatorial approaches to chiral selector methods. For example, high throughput screening methods of the invention can be used to screen libraries of peptides to identify those sequences with unique enantio-recognition properties; see, e.g., Chankvetadze (2001) *supra*. Thus, the invention provides chimeric monomers and polymers, including nanotubules, comprising libraries

of peptides. In one aspect, these peptide sequences are inserted into the sequence of a chimeric monomer and uniformly displayed on the nanotubule surface.

In one aspect, to apply evolution technologies to the development of chiral selectors, the invention provides a high throughput screen suitable for the identification of 5 protein variants that possess increased enantioselectivity. For example, Henriksson (1996) *supra*, have reported that the activity of cellobiohydrolase (CBH) from *Trichoderma reesei* is differentially inhibited by the (R)- and (S)-enantiomers of the beta-blockers propanolol and alprenolol. The *T. reesei* CBH has been demonstrated to be an effective chiral selector for beta-blockers and the chiral selectivity is consistent with the 10 inhibition data. Based on these results, the methods of the invention evolve the enantioselectivity of CBH using evolution strategies. In one aspect, a high throughput screen is used that measures enantiospecific inhibition of CBH activity.

In practicing the invention, a nucleic acid (e.g., a nucleic acid encoding a chimeric polypeptide of the invention) can be altered by any means. For example, 15 random or stochastic methods, or, non-stochastic, or “directed evolution,” methods, see, e.g., U.S. Patent No. 6,361,974. Methods for random mutation of genes are well known in the art, see, e.g., U.S. Patent No. 5,830,696. For example, mutagens can be used to randomly mutate a gene. Mutagens include, e.g., ultraviolet light or gamma irradiation, or a chemical mutagen, e.g., mitomycin, nitrous acid, photoactivated psoralens, alone or 20 in combination, to induce DNA breaks amenable to repair by recombination. Other chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. Other mutagens are analogues of nucleotide precursors, e.g., nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine. These agents can be added to a PCR reaction in place of the nucleotide precursor thereby mutating the sequence. 25 Intercalating agents such as proflavine, acriflavine, quinacrine and the like can also be used.

Any technique in molecular biology can be used, e.g., random PCR mutagenesis, see, e.g., Rice (1992) *Proc. Natl. Acad. Sci. USA* 89:5467-5471; or, 30 combinatorial multiple cassette mutagenesis, see, e.g., Crameri (1995) *Biotechniques* 18:194-196. Alternatively, nucleic acids, e.g., genes, can be reassembled after random, or “stochastic,” fragmentation, see, e.g., U.S. Patent Nos. 6,291,242; 6,287,862; 6,287,861; 5,955,358; 5,830,721; 5,824,514; 5,811,238; 5,605,793. In alternative aspects, modifications, additions or deletions are introduced by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis (GSSM), synthetic ligation reassembly (SLR), recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation, and/or a combination of these and other methods.

The following publications describe a variety of recursive recombination procedures and/or methods which can be incorporated into the methods of the invention:

Stemmer (1999) "Molecular breeding of viruses for targeting and other clinical properties" *Tumor Targeting* 4:1-4; Ness (1999) *Nature Biotechnology* 17:893-896;

Chang (1999) "Evolution of a cytokine using DNA family shuffling" *Nature Biotechnology* 17:793-797; Minshull (1999) "Protein evolution by molecular breeding" *Current Opinion in Chemical Biology* 3:284-290; Christians (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Crameri (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Crameri (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," *Nature Biotechnology* 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" *Nature Medicine* 2:100-103; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" *Journal of Molecular Biology* 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: *The Encyclopedia of Molecular Biology*. VCH Publishers, New York. pp.447-457;

Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence

Space" Bio/Technology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." Proc. Natl. Acad. Sci. USA 91:10747-10751.

5 Mutational methods of generating variant sequences in practicing the methods of the invention include, for example, site-directed mutagenesis (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol. Biol. 57:369-374; Smith (1985) "In vitro 10 mutagenesis" Ann. Rev. Genet. 19:423-462; Botstein & Shortle (1985) "Strategies and applications of in vitro mutagenesis" Science 229:1193-1201; Carter (1986) "Site-directed mutagenesis" Biochem. J. 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates 15 (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Methods in Enzymol. 154, 367-382; and Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" Science 242:240-245); oligonucleotide-directed mutagenesis (Methods in 20 Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" Nucleic Acids Res. 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" Methods in Enzymol. 25 100:468-500; and Zoller & Smith (1987) Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" Methods in Enzymol. 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" Nucl. Acids Res. 13: 8749-8764; Taylor et al. (1985) "The rapid 30 generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucl. Acids Res. 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers et al. (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-

directed mutagenesis" *Nucl. Acids Res.* 16:791-802; and Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814); mutagenesis using gapped duplex DNA (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456; Kramer & Fritz (1987) *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer et al. (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" *Nucl. Acids Res.* 16: 7207; and Fritz et al. (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" *Nucl. Acids Res.* 16: 6987-6999).

Additional protocols for generating variant sequences in practicing the methods of the invention include point mismatch repair (Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" *Methods in Enzymol.* 154: 382-403), deletion mutagenesis (Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115), restriction-selection and restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" *Phil. Trans. R. Soc. Lond. A* 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" *Nucl. Acids Res.* 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" *Gene* 34:315-323; and Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" *Nucl. Acids Res.* 13: 3305-3316), double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" *Current Opinion in Biotechnology* 4:450-455. "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis" *Proc. Natl. Acad. Sci. USA*, 83:7177-7181).

Additional details on many of the above methods can be found in Methods in

Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Additional protocols for generating variant sequences in practicing the methods of the invention include those discussed in U.S. Patent Nos. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for In Vitro Recombination;" U.S. Pat. No. 5,811,238 to Stemmer et al. (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer et al. (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, et al. (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 to Minshull, et al. (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 by Stemmer and Crameri "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen et al. "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen et al. "Antigen Library Immunization;" WO 99/41369 by Punnonen et al. "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt et al., "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/27230 by Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection," WO 00/00632, "Methods for Generating Highly Diverse Libraries," WO 00/09679, "Methods for Obtaining in Vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences," WO 98/42832 by Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers," WO 99/29902 by Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences," WO 98/41653 by Vind, "An in

"Vitro Method for Construction of a DNA Library," WO 98/41622 by Borchert et al., "Method for Constructing a Library Using DNA Shuffling," and WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination."

Additional protocols for generating variant sequences in practicing the methods of the invention are described in U.S. Patent application serial no. (USSN) 09/407,800, "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed Sep. 28, 1999; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" by del Cardayre et al., United States Patent No. 6,379,964; "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., United States Patent Nos. 6,319,714; 6,368,861; 6,376,246; 6,423,542; 6,426,224 and PCT/US00/01203; "USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., United States Patent No. 6,436,675; "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed Jan. 18, 2000, (PCT/US00/01202) and, e.g. "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., filed Jul. 18, 2000 (U.S. Ser. No. 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" by Selifonov and Stemmer, filed Jan. 18, 2000 (PCT/US00/01138); and "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" by Affholter, filed Sep. 6, 2000 (U.S. Ser. No. 09/656,549); and United States Patent Nos. 6,177,263; 6,153,410.

Non-stochastic, or "directed evolution," methods

Exemplary protocols for generating variant sequences (e.g., modified sequences encoding chimeric polypeptides of the invention) in practicing the methods of the invention include non-stochastic, or "directed evolution," methods, such as, e.g., saturation mutagenesis (GSSM), synthetic ligation reassembly (SLR), or a combination thereof. These methods can be used to modify the nucleic acids to generate chimeric polypeptides with new or altered properties (e.g., chiral selection activity under high or low acidic or alkaline conditions, high or low temperatures, high or low salt conditions and the like; different substrate affinity; enantioselective activity; modified antibody binding activity, etc.). Polypeptides encoded by the modified nucleic acids can be

screened for an activity before testing for proteolytic or other activity. Any testing modality or protocol can be used, e.g., using a capillary array platform. See, e.g., U.S. Patent Nos. 6,361,974; 6,280,926; 5,939,250.

Saturation mutagenesis, or, GSSM

5 In one aspect of the invention, non-stochastic gene modification, a “directed evolution process,” is used to generate modified sequences encoding chimeric polypeptides of the invention with new or altered properties. Variations of this method have been termed “gene site-saturation mutagenesis,” “site-saturation mutagenesis,” “saturation mutagenesis” or simply “GSSM.” It can be used in combination with other 10 mutagenization processes. See, e.g., U.S. Patent Nos. 6,171,820; 6,238,884. In one aspect, GSSM comprises providing a template polynucleotide and a plurality of oligonucleotides, wherein each oligonucleotide comprises a sequence homologous to the template polynucleotide, thereby targeting a specific sequence of the template polynucleotide, and a sequence that is a variant of the homologous gene; generating 15 progeny polynucleotides comprising non-stochastic sequence variations by replicating the template polynucleotide with the oligonucleotides, thereby generating polynucleotides comprising homologous gene sequence variations.

20 In one aspect, codon primers containing a degenerate N,N,G/T sequence are used to introduce point mutations into a polynucleotide, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is 25 represented at each amino acid position, e.g., an amino acid residue in an enzyme active site or ligand binding site targeted to be modified. These oligonucleotides can comprise a contiguous first homologous sequence, a degenerate N,N,G/T sequence, and, optionally, a second homologous sequence. The downstream progeny translational products from the use of such oligonucleotides include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,G/T sequence includes codons for all 20 amino acids. In one aspect, one such degenerate oligonucleotide (comprised of, e.g., one degenerate N,N,G/T cassette) is used for subjecting each original codon in a parental polynucleotide template to a full range of codon substitutions. In 30 another aspect, at least two degenerate cassettes are used – either in the same oligonucleotide or not, for subjecting at least two original codons in a parental polynucleotide template to a full range of codon substitutions. For example, more than one N,N,G/T sequence can be contained in one oligonucleotide to introduce amino acid

mutations at more than one site. This plurality of N,N,G/T sequences can be directly contiguous, or separated by one or more additional nucleotide sequence(s). In another aspect, oligonucleotides serviceable for introducing additions and deletions can be used either alone or in combination with the codons containing an N,N,G/T sequence, to 5 introduce any combination or permutation of amino acid additions, deletions, and/or substitutions.

In one aspect, simultaneous mutagenesis of two or more contiguous amino acid positions is done using an oligonucleotide that contains contiguous N,N,G/T triplets, i.e. a degenerate (N,N,G/T)_n sequence. In another aspect, degenerate cassettes having 10 less degeneracy than the N,N,G/T sequence are used. For example, it may be desirable in some instances to use (e.g. in an oligonucleotide) a degenerate triplet sequence comprised of only one N, where said N can be in the first second or third position of the triplet. Any other bases including any combinations and permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some 15 instances to use (e.g. in an oligo) a degenerate N,N,N triplet sequence.

In one aspect, use of degenerate triplets (e.g., N,N,G/T triplets) allows for systematic and easy generation of a full range of possible natural amino acids (for a total of 20 amino acids) into each and every amino acid position in a polypeptide (in alternative aspects, the methods also include generation of less than all possible 20 substitutions per amino acid residue, or codon, position). For example, for a 100 amino acid polypeptide, 2000 distinct species (i.e. 20 possible amino acids per position X 100 amino acid positions) can be generated. Through the use of an oligonucleotide or set of oligonucleotides containing a degenerate N,N,G/T triplet, 32 individual sequences can code for all 20 possible natural amino acids. Thus, in a reaction vessel in which a 25 parental polynucleotide sequence is subjected to saturation mutagenesis using at least one such oligonucleotide, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligonucleotide in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel. Nondegenerate oligonucleotides can optionally be used in combination with degenerate 30 primers disclosed; for example, nondegenerate oligonucleotides can be used to generate specific point mutations in a working polynucleotide. This provides one means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes, and point mutations that cause the generation of stop codons and the corresponding expression of polypeptide fragments.

In one aspect, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide molecules such that all 20 natural amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide (other aspects use less than all 20 natural combinations). The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (e.g. cloned into a suitable host, e.g., *E. coli* host, using, e.g., an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in property (when compared to the parental polypeptide, such as increased proteolytic activity under alkaline or acidic conditions), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

In one aspect, upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid, and each of two favorable changes) and 3 positions. Thus, there are $3 \times 3 \times 3$ or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (i.e. 2 at each of three positions) and no change at any position.

In another aspect, site-saturation mutagenesis can be used together with another stochastic or non-stochastic means to vary sequence, e.g., synthetic ligation reassembly (see below), shuffling, chimerization, recombination and other mutagenizing processes and mutagenizing agents. This invention provides for the use of any mutagenizing process(es), including saturation mutagenesis, in an iterative manner.

Synthetic Ligation Reassembly (SLR)

In practicing the methods of the invention a non-stochastic gene modification system termed “synthetic ligation reassembly,” or simply “SLR,” a “directed evolution process,” can be used to generate modified sequences encoding chimeric polypeptides of the invention with new or altered properties. SLR is a method of ligating oligonucleotide fragments together non-stochastically. This method differs from

stochastic oligonucleotide shuffling in that the nucleic acid building blocks are not shuffled, concatenated or chimerized randomly, but rather are assembled non-stochastically. See, e.g., U.S. Patent Application Serial No. (USSN) 09/332,835 entitled “Synthetic Ligation Reassembly in Directed Evolution” and filed on June 14, 1999 (“USSN 09/332,835”). In one aspect, SLR comprises the following steps: (a) providing a template polynucleotide, wherein the template polynucleotide comprises sequence encoding a homologous gene; (b) providing a plurality of building block polynucleotides, wherein the building block polynucleotides are designed to cross-over reassemble with the template polynucleotide at a predetermined sequence, and a building block 5 polynucleotide comprises a sequence that is a variant of the homologous gene and a sequence homologous to the template polynucleotide flanking the variant sequence; (c) combining a building block polynucleotide with a template polynucleotide such that the building block polynucleotide cross-over reassembles with the template polynucleotide to generate polynucleotides comprising homologous gene sequence variations.

10

15 SLR does not depend on the presence of high levels of homology between polynucleotides to be rearranged. Thus, this method can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over 10100 different chimeras. SLR can be used to generate libraries comprised of over 101000 different progeny chimeras. Thus, aspects of the present invention include non-stochastic methods 20 of producing a set of finalized chimeric nucleic acid molecule shaving an overall assembly order that is chosen by design. This method includes the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

25 The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be “serviceable” for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends. If more than one assembly step is to be used, then the 30 overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). In one aspect, the annealed building pieces are treated with an enzyme, such as a ligase (e.g. T4 DNA ligase), to achieve covalent bonding of the building pieces.

In one aspect, the design of the oligonucleotide building blocks is obtained by analyzing a set of progenitor nucleic acid sequence templates that serve as a basis for producing a progeny set of finalized chimeric polynucleotides. These parental oligonucleotide templates thus serve as a source of sequence information that aids in the 5 design of the nucleic acid building blocks that are to be mutagenized, e.g., chimerized or shuffled. In one aspect of this method, the sequences of a plurality of parental nucleic acid templates are aligned in order to select one or more demarcation points. The demarcation points can be located at an area of homology, and are comprised of one or more nucleotides. These demarcation points are preferably shared by at least two of the 10 progenitor templates. The demarcation points can thereby be used to delineate the boundaries of oligonucleotide building blocks to be generated in order to rearrange the parental polynucleotides. The demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the final chimeric progeny molecules. A demarcation point can be an area of homology 15 (comprised of at least one homologous nucleotide base) shared by at least two parental polynucleotide sequences. Alternatively, a demarcation point can be an area of homology that is shared by at least half of the parental polynucleotide sequences, or, it can be an area of homology that is shared by at least two thirds of the parental polynucleotide sequences. Even more preferably a serviceable demarcation points is an area of 20 homology that is shared by at least three fourths of the parental polynucleotide sequences, or, it can be shared by at almost all of the parental polynucleotide sequences. In one aspect, a demarcation point is an area of homology that is shared by all of the parental polynucleotide sequences.

In one aspect, a ligation reassembly process is performed exhaustively in 25 order to generate an exhaustive library of progeny chimeric polynucleotides. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, in another aspect, the assembly order (i.e. the order of assembly of each building block in the 5' to 3' sequence of each finalized chimeric nucleic acid) in each combination is by 30 design (or non-stochastic) as described above. Because of the non-stochastic nature of this invention, the possibility of unwanted side products is greatly reduced.

In another aspect, the ligation reassembly method is performed systematically. For example, the method is performed in order to generate a systematically compartmentalized library of progeny molecules, with compartments that

can be screened systematically, e.g. one by one. In other words this invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, a design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure to be performed. Thus, these methods allow a potentially very large number of progeny molecules to be examined systematically in smaller groups. Because of its ability to perform chimerizations in a manner that is highly flexible yet exhaustive and systematic as well, particularly when there is a low level of homology among the progenitor molecules, these methods provide for the generation of a library (or set) comprised of a large number of progeny molecules. Because of the non-stochastic nature of the instant ligation reassembly invention, the progeny molecules generated preferably comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. The saturation mutagenesis and optimized directed evolution methods also can be used to generate different progeny molecular species. It is appreciated that the invention provides freedom of choice and control regarding the selection of demarcation points, the size and number of the nucleic acid building blocks, and the size and design of the couplings. It is appreciated, furthermore, that the requirement for intermolecular homology is highly relaxed for the operability of this invention. In fact, demarcation points can even be chosen in areas of little or no intermolecular homology. For example, because of codon wobble, i.e. the degeneracy of codons, nucleotide substitutions can be introduced into nucleic acid building blocks without altering the amino acid originally encoded in the corresponding progenitor template. Alternatively, a codon can be altered such that the coding for an originally amino acid is altered. This invention provides that such substitutions can be introduced into the nucleic acid building block in order to increase the incidence of intermolecular homologous demarcation points and thus to allow an increased number of couplings to be achieved among the building blocks, which in turn allows a greater number of progeny chimeric molecules to be generated.

In another aspect, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (e.g., one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that can later be optionally removed in an in vitro process (e.g. by mutagenesis) or in an in vivo process (e.g. by utilizing the gene splicing ability of a host organism). It is

appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.

5 In one aspect, a nucleic acid building block is used to introduce an intron. Thus, functional introns are introduced into a man-made gene manufactured according to the methods described herein. The artificially introduced intron(s) can be functional in a host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing.

Optimized Directed Evolution System

10 In practicing the methods of the invention a non-stochastic gene modification system termed “optimized directed evolution system” can be used to generate modified sequences encoding chimeric polypeptides of the invention with new or altered properties. Optimized directed evolution is directed to the use of repeated cycles of reductive reassortment, recombination and selection that allow for the directed 15 molecular evolution of nucleic acids through recombination. Optimized directed evolution allows generation of a large population of evolved chimeric sequences, wherein the generated population is significantly enriched for sequences that have a predetermined number of crossover events.

20 A crossover event is a point in a chimeric sequence where a shift in sequence occurs from one parental variant to another parental variant. Such a point is normally at the juncture of where oligonucleotides from two parents are ligated together to form a single sequence. This method allows calculation of the correct concentrations of oligonucleotide sequences so that the final chimeric population of sequences is enriched for the chosen number of crossover events. This provides more control over 25 choosing chimeric variants having a predetermined number of crossover events.

30 In addition, this method provides a convenient means for exploring a tremendous amount of the possible protein variant space in comparison to other systems. Previously, if one generated, for example, 10^{13} chimeric molecules during a reaction, it would be extremely difficult to test such a high number of chimeric variants for a particular activity. Moreover, a significant portion of the progeny population would have a very high number of crossover events which resulted in proteins that were less likely to have increased levels of a particular activity. By using these methods, the population of chimerics molecules can be enriched for those variants that have a particular number of

crossover events. Thus, although one can still generate 10^{13} chimeric molecules during a reaction, each of the molecules chosen for further analysis most likely has, for example, only three crossover events. Because the resulting progeny population can be skewed to have a predetermined number of crossover events, the boundaries on the functional variety between the chimeric molecules is reduced. This provides a more manageable number of variables when calculating which oligonucleotide from the original parental polynucleotides might be responsible for affecting a particular trait.

One method for creating a chimeric progeny polynucleotide sequence is to create oligonucleotides corresponding to fragments or portions of each parental sequence. 5 Each oligonucleotide preferably includes a unique region of overlap so that mixing the oligonucleotides together results in a new variant that has each oligonucleotide fragment assembled in the correct order. Additional information can also be found, e.g., in USSN 10 09/332,835; U.S. Patent No. 6,361,974. The number of oligonucleotides generated for each parental variant bears a relationship to the total number of resulting crossovers in the 15 chimeric molecule that is ultimately created. For example, three parental nucleotide sequence variants might be provided to undergo a ligation reaction in order to find a chimeric variant having, for example, greater activity at high temperature. As one example, a set of 50 oligonucleotide sequences can be generated corresponding to each portions of each parental variant. Accordingly, during the ligation reassembly process 20 there could be up to 50 crossover events within each of the chimeric sequences. The probability that each of the generated chimeric polynucleotides will contain oligonucleotides from each parental variant in alternating order is very low. If each 25 oligonucleotide fragment is present in the ligation reaction in the same molar quantity it is likely that in some positions oligonucleotides from the same parental polynucleotide will ligate next to one another and thus not result in a crossover event. If the concentration of each oligonucleotide from each parent is kept constant during any ligation step in this example, there is a 1/3 chance (assuming 3 parents) that an oligonucleotide from the same 30 parental variant will ligate within the chimeric sequence and produce no crossover.

Accordingly, a probability density function (PDF) can be determined to predict the population of crossover events that are likely to occur during each step in a ligation reaction given a set number of parental variants, a number of oligonucleotides corresponding to each variant, and the concentrations of each variant during each step in the ligation reaction. The statistics and mathematics behind determining the PDF is described below. By utilizing these methods, one can calculate such a probability density

function, and thus enrich the chimeric progeny population for a predetermined number of crossover events resulting from a particular ligation reaction. Moreover, a target number of crossover events can be predetermined, and the system then programmed to calculate the starting quantities of each parental oligonucleotide during each step in the ligation reaction to result in a probability density function that centers on the predetermined number of crossover events. These methods are directed to the use of repeated cycles of reductive reassortment, recombination and selection that allow for the directed molecular evolution of a nucleic acid encoding a polypeptide through recombination. This system allows generation of a large population of evolved chimeric sequences, wherein the generated population is significantly enriched for sequences that have a predetermined number of crossover events. A crossover event is a point in a chimeric sequence where a shift in sequence occurs from one parental variant to another parental variant. Such a point is normally at the juncture of where oligonucleotides from two parents are ligated together to form a single sequence. The method allows calculation of the correct concentrations of oligonucleotide sequences so that the final chimeric population of sequences is enriched for the chosen number of crossover events. This provides more control over choosing chimeric variants having a predetermined number of crossover events.

In addition, these methods provide a convenient means for exploring a tremendous amount of the possible protein variant space in comparison to other systems. By using the methods described herein, the population of chimeric molecules can be enriched for those variants that have a particular number of crossover events. Thus, although one can still generate 10^{13} chimeric molecules during a reaction, each of the molecules chosen for further analysis most likely has, for example, only three crossover events. Because the resulting progeny population can be skewed to have a predetermined number of crossover events, the boundaries on the functional variety between the chimeric molecules is reduced. This provides a more manageable number of variables when calculating which oligonucleotide from the original parental polynucleotides might be responsible for affecting a particular trait.

In one aspect, the method creates a chimeric progeny polynucleotide sequence by creating oligonucleotides corresponding to fragments or portions of each parental sequence. Each oligonucleotide preferably includes a unique region of overlap so that mixing the oligonucleotides together results in a new variant that has each oligonucleotide fragment assembled in the correct order. See also USSN 09/332,835.

The number of oligonucleotides generated for each parental variant bears a relationship to the total number of resulting crossovers in the chimeric molecule that is ultimately created. For example, three parental nucleotide sequence variants might be provided to undergo a ligation reaction in order to find a chimeric variant having, for 5 example, greater activity at high temperature. As one example, a set of 50 oligonucleotide sequences can be generated corresponding to each portions of each parental variant. Accordingly, during the ligation reassembly process there could be up to 50 crossover events within each of the chimeric sequences. The probability that each of 10 the generated chimeric polynucleotides will contain oligonucleotides from each parental variant in alternating order is very low. If each oligonucleotide fragment is present in the ligation reaction in the same molar quantity it is likely that in some positions oligonucleotides from the same parental polynucleotide will ligate next to one another and thus not result in a crossover event. If the concentration of each oligonucleotide from 15 each parent is kept constant during any ligation step in this example, there is a 1/3 chance (assuming 3 parents) that an oligonucleotide from the same parental variant will ligate within the chimeric sequence and produce no crossover.

Accordingly, a probability density function (PDF) can be determined to predict the population of crossover events that are likely to occur during each step in a ligation reaction given a set number of parental variants, a number of oligonucleotides 20 corresponding to each variant, and the concentrations of each variant during each step in the ligation reaction. The statistics and mathematics behind determining the PDF is described below. One can calculate such a probability density function, and thus enrich the chimeric progeny population for a predetermined number of crossover events resulting from a particular ligation reaction. Moreover, a target number of crossover 25 events can be predetermined, and the system then programmed to calculate the starting quantities of each parental oligonucleotide during each step in the ligation reaction to result in a probability density function that centers on the predetermined number of crossover events.

Iterative Processes

In practicing the invention, these processes can be iteratively repeated. 30 For example a nucleic acid (or, the nucleic acid) responsible for an altered phenotype of a chimeric polypeptide of the invention is identified, re-isolated, again modified, re-tested for activity using the methods of the invention. This process can be iteratively repeated

until a desired phenotype is engineered. For example, an entire biochemical anabolic or catabolic pathway can be engineered into a cell, including proteolytic activity.

Similarly, if it is determined that a particular oligonucleotide has no affect at all on the desired trait, it can be removed as a variable by synthesizing larger parental oligonucleotides that include the sequence to be removed. Since incorporating the sequence within a larger sequence prevents any crossover events, there will no longer be any variation of this sequence in the progeny polynucleotides. This iterative practice of determining which oligonucleotides are most related to the desired trait, and which are unrelated, allows more efficient exploration all of the possible protein variants that might provide a particular trait or activity.

Producing sequence variants

In practicing the methods of the invention nucleic acid variants can be generated using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives may be created using chemical synthesis or modification procedures. Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids which encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. These nucleotide differences can result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants may be created using error prone PCR. In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Error prone PCR is described, e.g., in Leung, D.W., et al., *Technique*, 1:11-15, 1989 and Caldwell, R. C. & Joyce G.F., *PCR Methods Applic.*, 2:28-33, 1992. Briefly, in such procedures, nucleic acids to be mutagenized are mixed with PCR primers, reaction buffer, MgCl₂, MnCl₂, Taq polymerase and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction may be performed using 20 fmoles of

nucleic acid to be mutagenized, 30 pmole of each PCR primer, a reaction buffer comprising 50 mM KCl, 10 mM Tris HCl (pH 8.3) and 0.01% gelatin, 7 mM MgCl₂, 0.5 mM MnCl₂, 5 units of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM dTTP. PCR may be performed for 30 cycles of 94°C for 1 min, 45°C for 1 min, 5 and 72°C for 1 min. However, it will be appreciated that these parameters may be varied as appropriate. The mutagenized nucleic acids are cloned into an appropriate vector and the activities of the polypeptides encoded by the mutagenized nucleic acids is evaluated.

10 Variants may also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described, e.g., in Reidhaar-Olson (1988) *Science* 241:53-57. Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized. Clones containing the mutagenized DNA are recovered and the activities of the polypeptides they encode are assessed.

15 Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in, e.g., U.S. Patent No. 5,965,408.

20 Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different but highly related DNA sequence *in vitro*, as a result of random fragmentation of the DNA molecule based on sequence homology, followed by fixation of the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described, e.g., in Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751. Briefly, in such procedures a plurality of nucleic acids to be recombined are digested with DNase to generate fragments having an average size of 50-200 nucleotides. Fragments of the desired average size are purified and resuspended in a PCR mixture. PCR is conducted under conditions which facilitate recombination between the nucleic acid fragments. For example, PCR may be performed by resuspending the purified fragments at a concentration of 10-30 ng:/l in a solution of 0.2 mM of each dNTP, 2.2 mM MgCl₂, 50 mM KCL, 10 mM Tris HCl, pH 9.0, and 0.1% Triton X-100. 2.5 units of Taq polymerase per 100:l of reaction mixture is added and PCR is performed using the following regime: 94°C for 60 seconds, 94°C for 30 seconds, 50-55°C for 30 seconds, 72°C for 30 seconds

(30-45 times) and 72°C for 5 minutes. However, it will be appreciated that these parameters may be varied as appropriate. In some aspects, oligonucleotides may be included in the PCR reactions. In other aspects, the Klenow fragment of DNA polymerase I may be used in a first set of PCR reactions and Taq polymerase may be used in a subsequent set of PCR reactions. Recombinant sequences are isolated and the activities of the polypeptides they encode are assessed.

5 Variants may also be created by *in vivo* mutagenesis. In some aspects, random mutations in a sequence of interest are generated by propagating the sequence of interest in a bacterial strain, such as an *E. coli* strain, which carries mutations in one or 10 more of the DNA repair pathways. Such "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for *in vivo* mutagenesis are described, e.g., in PCT Publication No. WO 91/16427.

15 Variants may also be generated using cassette mutagenesis. In cassette mutagenesis a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

20 Recursive ensemble mutagenesis may also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described, e.g., in Arkin (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815.

25 In some aspects, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is described, e.g., in Delegrave (1993) Biotechnology Res. 11:1548-1552. Random and 30 site-directed mutagenesis are described, e.g., in Arnold (1993) Current Opinion in Biotechnology 4:450-455.

In some aspects, the variants are created using shuffling procedures wherein portions of a plurality of nucleic acids which encode distinct polypeptides are

fused together to create chimeric nucleic acid sequences which encode chimeric polypeptides as described in, e.g., U.S. Patent Nos. 5,965,408; 5,939,250.

Optimizing codons to achieve high levels of protein expression in host cells

In one aspect of the invention, nucleic acids are mutated to modify codon usage. In one aspect, methods of the invention comprise modifying codons in a nucleic acid encoding a modified sequence encoding a chimeric polypeptide of the invention to increase or decrease its expression in a host cell, e.g., a bacterial, insect, mammalian, yeast or plant cell. The method can comprise identifying a “non-preferred” or a “less preferred” codon in protein-encoding nucleic acid and replacing one or more of these non-preferred or less preferred codons with a “preferred codon” encoding the same amino acid as the replaced codon and at least one non-preferred or less preferred codon in the nucleic acid has been replaced by a preferred codon encoding the same amino acid. A preferred codon is a codon over-represented in coding sequences in genes in the host cell and a non-preferred or less preferred codon is a codon under-represented in coding sequences in genes in the host cell.

Methodologies and Devices

In practicing the invention, a variety of apparatus and methodologies can be used, e.g., using the chimeric monomers and polymers for chiral selection, to determine the efficiency of the chiral separation from a racemic mixture, as biosynthetic pathways, as selection scaffoldings, to screen for variant chimeric polypeptides, to determine the extent of nanotubule formation, and the like.

Capillary Arrays

Capillary arrays, such as the GIGAMATRIX™, Diversa Corporation, San Diego, CA, can be used to practice the invention. Nucleic acids or polypeptides (the chimeric monomers and polymers of the invention) or other compositions (e.g., substrates or co-factors for using the nanotubule biosynthetic pathways of the invention, antibodies or other compounds for binding to chimeric monomers of the invention) can be immobilized to or applied to an array, including capillary arrays. Arrays can be used in the chiral selection methods of the invention. Capillary arrays can provide a system for holding and screening samples, monomers of the invention, chiral products selected by the methods of the invention, substrates and co-factors and products used in the biosynthetic pathway methods of the invention, and the like.

A sample apparatus can include a plurality of capillaries formed into an array of adjacent capillaries, wherein each capillary comprises at least one wall defining a lumen for retaining a sample. The apparatus can further include interstitial material disposed between adjacent capillaries in the array, and one or more reference indicia formed within of the interstitial material. A capillary for screening a sample, wherein the capillary is adapted for being bound in an array of capillaries, can include a first wall defining a lumen for retaining the sample, and a second wall formed of a filtering material, for filtering excitation energy provided to the lumen to excite the sample.

A polypeptide or other composition can be introduced into a first component into at least a portion of a capillary of a capillary array. Each capillary of the capillary array can comprise at least one wall defining a lumen for retaining the first component. An air bubble can be introduced into the capillary behind the first component. A second component can be introduced into the capillary, wherein the second component is separated from the first component by the air bubble. A sample of interest can be introduced as a first liquid labeled with a detectable particle into a capillary of a capillary array, wherein each capillary of the capillary array comprises at least one wall defining a lumen for retaining the first liquid and the detectable particle, and wherein the at least one wall is coated with a binding material for binding the detectable particle to the at least one wall. The method can further include removing the first liquid from the capillary tube, wherein the bound detectable particle is maintained within the capillary, and introducing a second liquid into the capillary tube.

The capillary array can include a plurality of individual capillaries comprising at least one outer wall defining a lumen. The outer wall of the capillary can be one or more walls fused together. Similarly, the wall can define a lumen that is cylindrical, square, hexagonal or any other geometric shape so long as the walls form a lumen for retention of a liquid or sample. The capillaries of the capillary array can be held together in close proximity to form a planar structure. The capillaries can be bound together, by being fused (e.g., where the capillaries are made of glass), glued, bonded, or clamped side-by-side. The capillary array can be formed of any number of individual capillaries, for example, a range from 100 to 4,000,000 capillaries. A capillary array can form a micro titer plate having about 100,000 or more individual capillaries bound together.

Arrays, or "Biochips"

In practicing the invention polypeptides of the invention can be immobilized to or applied to an array. Arrays can be used to practice the methods of the invention, e.g., chiral selection from a racemic mixture. Polypeptide arrays" can be used 5 to simultaneously quantify or select for a plurality of proteins. The present invention can be practiced with any known "array," also referred to as a "microarray" or "DNA array" or "nucleic acid array" or "polypeptide array" or "antibody array" or "biochip," or variation thereof. Arrays are generically a plurality of "spots" or "target elements," each target element comprising a defined amount of one or more biological molecules 10 immobilized onto a defined area of a substrate surface for specific binding to a sample molecule. Any immobilization method can be used, e.g., immobilization upon an inert support such as diethylaminoethyl-cellulose, porous glass, chitin or cells. In practicing the methods of the invention, any known array and/or method of making and using arrays can be incorporated in whole or in part, or variations thereof, as described, for example, 15 in U.S. Patent Nos. 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; 5,434,049; see also, e.g., WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; 20 see also, e.g., Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. See also published U.S. patent applications Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

Antibodies

25 The compositions and methods of the invention can be practice using antibodies. For example, a heterologous polypeptide of a chimeric protein of the invention can be an antibody, e.g., a catalytic antibody for use in a biosynthetic pathway of the invention, or, an antibody that specifically binds to an enzyme, co-factor, substrate and the like for use in a biosynthetic pathway of the invention, or, an antibody that binds 30 to a chiral selection protein or peptide used in the methods of the invention. Antibodies also can be used in immunoprecipitation, staining, immunoaffinity columns, and the like, to, e.g., purify chiral selection products or products of the biosynthetic pathways of the invention.

Methods of doing assays, e.g., ELISAs, with polyclonal and monoclonal antibodies are known to those of skill in the art and described in the scientific and patent literature, see, e.g., Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2d ed.) Academic Press, New York, NY (1986); Kohler (1975) *Nature* 256:495; Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York. Antibodies also can be generated *in vitro*, e.g., using recombinant antibody binding site expressing phage display libraries, in addition to the traditional *in vivo* methods using animals. See, e.g., Hoogenboom (1997) *Trends Biotechnol.* 15:62-70; Katz (1997) *Annu. Rev. Biophys. Biomol. Struct.* 26:27-45.

The ability of proteins in a biological sample to bind to the antibody may be determined using any of a variety of procedures familiar to those skilled in the art. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. Particular assays include ELISA assays, sandwich assays, radioimmunoassays, and Western Blots.

20 Kits

The invention provides kits comprising materials for practicing the invention, including monomers and polymers, e.g., nanotubules, of the invention. The kits can comprise solutions for assembling the nanotubules of the invention. For example, the solutions can comprise various salts, as described herein. The kits also can contain instructional material teaching the methodologies and uses of the invention, as described herein.

EXAMPLES

Example 1: Isolating Recombinant Proteins From *E. coli*

30 The following example describes an exemplary assay to isolate recombinant "cannulae" or "can" proteins from *E. coli*.

All exemplary assays in this example used:

Low salt buffer: 80 mM NaCl, 50 mM Tris/ HCl (pH 7.5), 9% glycerol

High salt buffer: 1.2 M NaCl, 50 mM Tris/ HCl (pH 7.5), 9% glycerol

Bicinchoninic Acid Test (BCA): The test was conducted according to the manufacturer's guide (Sigma, Deisenhofen). To this end, aliquots of protein samples (CanA, B, C) and of known BSA dilutions were mixed with 50 times the volume of a fresh BCA/CuSO₄ (50:1) solution, incubated at 60°C for 30 min. and measured in the spectrometer at 562 nm after cooling to RT. The protein concentrations were measured with the BSA calibration line.

5 a) CanA and CanB

One gram of recombinant *E. coli* with a particular sequence such as CanA or CanB expressed was absorbed in 4 ml low salt buffer. Cell lysis was conducted with a 10 French press (2 x at 20,000 psi, American Instrument Co., Silver Spring, USA). After pelletizing the cell fragments (Eppendorf centrifuge, 13,000 rpm, 5 min., RT), the protein solution was incubated at 80°C for 20 min. Then the denatured proteins were removed by centrifugation (as above). The supernatant was passed at 1 ml/min through a Q Sepharose column (1 x 12 cm = 9.4 ml, Pharmacia, Freiburg). The eluent containing CanA or CanB 15 was collected. The collected eluant was treated with leupeptin (1 µg/µl) and concentrated by a factor of 3 - 4 (based on the volume) in 4 - 8 hours in the MACROSEP™ centrifuge concentrators (Pall Filtron, Dreieich) with an exclusion limit of 5 kDa. After determining the protein concentration with the BCA test, the purified protein was shock frozen in liquid nitrogen in 100 - 200 µl aliquots and stored at -80°C. In each working step, a sample was 20 taken and analyzed on an SDS polyacrylamide gel.

b) CanC

The first step of isolating CanC is same as that of CanA and CanB (see example 21.a). However, during the second step, CanC was retained on the Q sepharose. After flushing the column with low salt buffer, CanC was eluted from the column with a 25 salt gradient (80 - 750 mM, in 60 ml) and collected by fractionation (1 ml each). Following analysis of the individual fractions on an SDS polyacrylamide gel, the CanC-containing fractions were combined and dialyzed against the low salt buffer at 4°C overnight. Finally the protein solution was eluted at 1 ml/min through a 1 ml RESOURCEQ™ column (Pharmacia, Freiburg). Then a salt gradient (80 - 750 mM, in 60 ml) was applied and 0.5 30 ml fractions were collected. After analysis of the same on an SDS polyacrylamide gel, the CanC-containing fractions were combined again and dialyzed against low salt buffer overnight. Following addition of leupeptin (1 µg/µl), the solution was concentrated by a factor of 7 (based on the volume) in 6 hours in the MICROSEP™ centrifuge concentrators (Pall Filtron, Dreieich) with an exclusion limit of 5 kDa.

Example 2: Production of a CanA polymer

The following example describes an exemplary protocol to produce a CanA polymer, including a chimeric polypeptide or a nanotubule of the invention.

a) 300L Fermentor Culture of Recombinant *E. Coli*.

5 A 300 L culture of recombinant *E. coli* BL21 (DE3) harboring expression plasmid pEX-CAN-A (produced by attaching sequence substantially identical to SEQ ID NO. 1 to a vector pET17b using a procedure described in Example 20) was grown in a HTE-Fermentor (Bioengineering, Wald, Switzerland) at 37°C under aeration (165 L air / min.) and stirring (400 rpm) with a doubling time of about 40 min. At an O.D. (600nm) of 10 0.80, production of Can A protein was induced by addition of 30 grams of IPTG. Cells were harvested 3 hours after the induction and after being cooled down to 4°C. Cell yield: 1,610 grams (wet weight).

b) Production of the polymer.

i. French Press: 250 g frozen cell mass of recombinant *E. coli* (stored at -60°C) were suspended in 600 ml buffer (Tris-HCL 50 mM, pH 7.5, containing 80 mM NaCl and 9% (v/v) glycerol). Final volume: 900 ml. Cells were broken down by a French Press (Aminco; 1 x 20,000 PSI). The viscosity of the solution was lowered by shearing the DNA using an Ultraturrax blender and by adding additional 400 ml buffer.

20 ii. Centrifugation: Particles were removed by centrifugation (Sorvall SS34 rotor; 19,000 rpm, 15 min.) and a clear supernatant (called “crude extract”) was obtained.

25 iii. Heat Precipitation: To precipitate the heat-sensitive protein, the crude extract was heated to 100°C for 1 min. For example, the crude extract (1,200 ml) was pumped through a 75 cm long plastic hose (inner diameter, 5 mm; 4.75 ml/min) immersed in a 100°C hot water-glycerol-bath (water: glycerol= 1: 1). The outlet end of the plastic hose was passed through an ice bath to cool down the solution in the hose before solution was finally collected using an Erlenmeyer flask.

iv. Centrifugation: The heat-treated crude extract was centrifuged for 25 min. at 9,000 rpm in Sorvall rotor GSA. The clear supernatant was collected.

30 v. Ammonium sulfate Precipitation: To the clear supernatant (840 ml), a 100% saturated ammonium sulfate solution (452 ml) was added at 4°C (final ammonium sulfate concentration: 35% saturation). After 2 hours at 4°C, the precipitate was collected by centrifugation (1 hour; 13,000 rpm; Sowall rotor GSA). The precipitate was then solubilized in a buffer solution (final volume 171 ml; 12,35 mg protein/ml; 2,112 mg total protein) to form a protein solution. Finally, the protein solution was dialyzed by Rapid

Dialysis against another buffer solution until its conductivity was the same as that of the buffer (3 hours).

vi. Polymerization: The dialyzed protein solution was diluted by addition of buffer to a final protein concentration of 6.5 mg/ml (final volume 325 ml). Then, under shaking in a 1L Erlenmeyer flask at 100°C (in a water bath), the diluted protein solution was rapidly heated to 80°C and then immediately transferred into a 500 ml screw-capped storage bottle. The storage bottle contained 3.32 ml (21.58 mg protein) of “Polymer Primers” (the “Polymer Primers” had been prepared before by 4 times French Press-shearing of a prefabricated Polymer suspension). Then, CaCl and MgCl (each at 20 mM final concentration) were added to the mixture and the closed bottle was stored in an 60°C water bath. After addition of these salts, the solution became immediately turbid, indicating rapid polymerization of the protein units. After 10 min polymerization, the formed Polymer fibers were sheared by ultraturracing the solution for 20 seconds in order to create additional polymer primers to speed up polymerization. Traces of silicone antifoam may be added before the ultraturracing to reduce foaming. Typically, after 10 min. polymerization at 80°C, Polymer or polymer fibers could be observed under an electron microscope. After 1 to 2 hours of polymerization, protein polymers could be completely removed from the solution by centrifugation (15 min., 20,000 rpm, Sorvall rotor SS34), indicating complete polymerization.

Yield of polymer: 2.1 grams (protein) from 250 grams (wet weight) of E.coli (about 1 g Polymer (dry weight)/119 g E.coli).

vii. Storage: Wet: At 4°C in a buffer containing 10 mM Na-Azide. Dry: Freeze-drying the polymer after the polymer being washed with an 1/10diluted buffer followed by centrifugation.

25 Example 3: Preparation of Lipid Coated Drug Delivery Complexes

The following example describes an exemplary protocol to prepare lipid coated drug delivery complexes of the invention, e.g., pharmaceutical compositions comprising CanA, e.g., the chimeric polypeptides or nanotubules of the invention.

To a solution containing 3mg/ml monomeric protein units (e.g. Can A: 182 amino acids: MW = 19,830 daltons, having a sequence of SED ID NO. 2), a desired amount of drug molecules, and a sufficient amount of electrically neutral lipids, millimolar calcium and magnesium cations are added to form a mixture. The mixture is kept at ambient condition for a sufficient amount time until liposomes form. Thereafter, gel filtration

chromatography is carried out on the mixture and the liposomes contained in the mixture are size fractionated. The desired fractions of the liposomes are then heated to 50°C in the presence of millimolar amounts of calcium and magnesium cations to initiate the polymerization of the monomeric polypeptide units within each liposome. The 5 polymerization results in the extreme deformation of the liposomes and produces sealed lipid tubules containing the drug molecules.

Example 3: Preparation of Nanotubules

The following example describes an exemplary protocol to prepare nanotubules of the invention.

10 Through genetic isolation techniques the inventors isolated cannulae-producing genes and cloned them using *E. coli*. *canA*, *canB*, and *canC* genes were isolated and artificially grown, and reproduced. These cannulae protein subunits are the monomers which undergo polymerization. In order to see the monomer under a microscope, Green Fluorescent Protein (GFP) were added to the A, B, and C terminal 15 ends. Under some conditions the GFP proteins which were added acted to stabilize the polymers (e.g., the nanotubules) and allowed visualization of the polymerization reaction of cannulae forming monomer.

20 The fluorescent protein also was fused to the monomer to generate a fluorescent nanotube. These *canA*-GFP and GFP-*canA* fusion proteins could stabilize the protein for assembly to form a polymer, without the use of *canB* or *canC*. The effects of varying salt conditions for polymerization of these new proteins were unknown prior to these experiments.

25 The GFP protein was initially isolated from the *Aequorea victoria* is a reporter molecule and fluoresces green when exposed to ultraviolet light. The stability of the GFP protein is species independent and its frequent ability to fuse to other proteins without inhibiting any original function allowed it to be freely attached to the *canA* protein. Recent research in GFP proteins has been able to produce mutant GFP proteins that are viewable under the whole visible spectrum, resulting in live action footage 30 between living cells, and viewable fluorescence in light other than ultraviolet light. The GFP protein is also highly chemical and thermal stable, allowing experiments to be executed at temperatures at 80°C and higher, due to its very compact structure. Due to the chemical stability and species independency of the GFP protein, the protein *canA* with the GFP on the carboxy-terminal end was able to be clearly viewed under the microscope

without any interference with the original properties of the *canA* protein. However, without the addition of the GFP protein, the isolated *canA* protein did not efficiently polymerize due to instability of the protein, as it did in its native environment directly from the organism. Despite this finding, the properties and applications of the *canA* protein were unchanged from its native behavior by the addition of the GFP protein. This allowed the collection of decipherable data by use of the confocal microscope and viewing the green fluorescence from the fused GFP protein. In the GFP protein, the aromatic system of the chromophore determines the wavelength of fluorescence, i.e., the color of the fluorescence.

Due to the natural ability of the *Pyrodictium abyssi* organism to polymerize in the hydrothermal vents, a similar environment was generated to reproduce or enhance this reaction in the laboratory (the invention comprises methods for polymerizing chimeric proteins and nanotubules using such similar environments). In order to investigate this possibility, it was necessary to test the effects of different salts found in the deep sea vent environment on polymerization. One of a variety of choices of chemical salt catalysts, including copper sulfate, manganese sulfate, zinc sulfate, iron sulfate, magnesium sulfate, and calcium sulfate, lithium sulfate, and cobalt sulfate were also added as catalysts. Independent of the salt used, the polymerization reaction was almost instantaneous upon the addition of the appropriate chemical catalyst, although there were varying effects on the structure and length of the polymer being formed.

The initial monomer generated was a bright fluorescent green prior to the inclusion of chemical additives. After the polymerization reaction of the cannulae proteins, a cloudy precipitate formed that signified polymerization between the cannulae protein monomers. The reaction material was centrifuged to separate the polymerization product from excess reactants and was then examined under an Olympus Confocal microscope. Through the use of an argon (Ar) laser and reverse objective lenses, the fluorescent proteins were examined and photographed for study and comparison.

Capable of using over ten different laser types, the FluoView™500™ Confocal Microscope blocks all other light from entering the viewed specimen and gives a clear image of both fluorescence and nonfluorescent forms of the sample. In the case of the experiments described herein, the argon (Ar) laser was used to excite the GFP molecules of a fluorescent polymer of the invention on the slide in order to emit green light photons. The photons were viewed and captured on film to record the length of the individual strands of the polymers and also measure the dimensions of the strands. In a

confocal microscope the scanning and image capture are both acquired through the objective lens which is under the specimen. The confocal microscope provides both a clear and detailed image of the polymers. However, due to the depth of the sample there can be difficulty in deciphering the exact individual strands of polymer. Since it is not possible to three dimensionally rotate the photograph once taken, there can be difficulty confirming whether what was examined is a single strand of polymerized protein, or several polymers stacked on top each other, forming the appearance of a single multi-folded chain. However, adjusting the PMT, Offset, and Gain on the microscope settings, accurate and readable data on the newly polymerized monomer was obtained.

The results of these experiments demonstrated that the methods of the invention comprising use of “deep sea salts” or equivalents are very effective in synthesizing nanotubules. Before the “deep sea salt initiation” processes of the invention, previous experiments had used magnesium chloride and calcium chloride. With the use of these salts with the wild type monomer polymerization would sometimes take days to occur. After examination with the use of the confocal microscope with argon (Ar) laser scanning, minimal sized polymer chains were observed far from the desired longer interweaving protein chains. Similarly, these salts were poor for forming polymers with the GFP *canA* protein. However, the processes of the invention (i.e., adding seawater salts, seawater solutions as described herein, or equivalent) were effective in generating protein chains and nanotubules. With salts such as manganese sulfate, the polymer chains were more than triple their original length. Thus, the processes of the invention are used to generate bionanotubes and compositions comprising them.

In some experiments the addition of copper sulfate not only did the monomer not polymerize, the addition of copper sulfate salt also stopped the GFP fluorescence. This was particularly unexpected. The GFP protein is generally considered very chemically stable and unaffected by salts. This result suggested that it was not necessarily just the fluorescence of the GFP protein that was inhibited, but the entire protein was disassembled. Supporting evidence for this was that not only were the biological nanotubes not viewable with green fluorescence, they were nonexistent upon translucent light inspection when the experiment was performed with a copper sulfate salt initiator. The entire reaction was stopped seeming to indicate a degradation of unpolymerized GFP fusion protein. The slide under the confocal microscope appeared to be blank, however, when the light was switched to translucent light, the monomer was viewable, but no polymer strands existed.

With this observation, a new hypothesis was developed. It was theorized, that since the organism was found in an environment where all these salts existed together, that better polymerization would occur if the positive salts were mixed together without the inhibitory salts, such as copper sulfate that is found in these deep ocean sites.

5 After examining these results under the same conditions, and under a confocal microscope, using both Argon (Ar) and translucent light, it was found that with the addition of this mixture, these salts had varying results in different parts of the viewed microscope slide. This suggests that the mixture of the salts was not evenly distributed throughout the sample. In the mixed salt experiment slide, the results of the
10 polymerization yielded at best only slightly larger polymers than their corresponding single salt experiments. Interestingly, copper sulfate was not inhibitory in these mixed salt experiments. Due to these results it was theorized that if all of the salts, including those that have negative gain, were mixed together to mimic the condition that these organisms thrive in then favorable results might be found due to reproducing the
15 concentration balance found with the organism under original conditions. Thus, in one aspect, the invention provides processes comprising use of a solution comprising salts mixed together which are, in one aspect, the same or similar to the growth microenvironment of the organisms that naturally synthesize nanotubules comprising CanA, such as *Pyrodictium abyssi*.

20

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A chimeric polypeptide comprising at least a first domain comprising a cannulae polypeptide and a second domain comprising a heterologous polypeptide or peptide.

5

2. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide is inserted at the amino terminal end, the carboxy terminal end or internal to the cannulae polypeptide.

10

3. The chimeric polypeptide of claim 1, wherein the cannulae polypeptide comprises a protein having 70% sequence identity to SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, and is capable of assembling into a polymer, or, is capable of acting as a chiral selector.

15

4. The chimeric polypeptide of claim 3, wherein the cannulae polypeptide comprises a protein having 80% sequence identity to SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, and is capable of assembling into a polymer.

20

5. The chimeric polypeptide of claim 4, wherein the cannulae polypeptide comprises a protein having 90% sequence identity to SEQ ID NO:2; SEQ ID NO:4, or SEQ ID NO:6, and is capable of assembling into a polymer.

25

6. The chimeric polypeptide of claim 5, wherein the cannulae polypeptide comprises a protein having 95% sequence identity to SEQ ID NO:2; SEQ ID NO:4, or SEQ ID NO:6, and is capable of assembling into a polymer.

7. The chimeric polypeptide of claim 6, wherein the cannulae polypeptide comprises a protein having sequence as set forth in SEQ ID NO:2 SEQ ID NO:4, or SEQ ID NO:6.

30

8. The chimeric polypeptide of claim 1, wherein the cannulae polypeptide is capable of assembling into a polymer.

9. The chimeric polypeptide of claim 8, wherein the cannulae polypeptide is capable of assembling into a nanotubule.

10. The chimeric polypeptide of claim 9, wherein the heterologous 5 polypeptide or peptide is expressed in the inner lumen of the nanotubule.

11. The chimeric polypeptide of claim 9, wherein the heterologous polypeptide or peptide is expressed on the exterior of the nanotubule.

10 12. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide comprises a chiral selection motif.

13. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide comprises a receptor or a ligand.

15 14. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide comprises an enzyme.

20 15. The chimeric polypeptide of claim 14, wherein the heterologous polypeptide or peptide comprises an enzyme active site.

16. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide comprises an antigen or an antigen binding site.

25 17. The chimeric polypeptide of claim 1, wherein the heterologous polypeptide or peptide comprises a green fluorescent protein, an alpha-galactosidase or a chloramphenicol acetyltransferase.

30 18. The chimeric polypeptide of claim 1, wherein the chimeric polypeptide is a recombinant protein.

19. The chimeric polypeptide of claim 1, wherein at least one subsequence of the cannulae polypeptide has been removed.

20. The chimeric polypeptide of claim 19, wherein the heterologous polypeptide or peptide is inserted into the cannulae polypeptide at the site the subsequence was removed.

5 21. The chimeric polypeptide of claim 19, wherein the cannulae polypeptide is a CanA polypeptide and the removed subsequence is a 14 residue motif consisting of residue 123 to residue 136 of SEQ ID NO:2 (PDKTGYTNTSIWVP), or, a 17 residue motif located at amino acid residue 123 to residue 139 of SEQ ID NO:2 (PDKTGYTNTSIWVPGE^P).

10

22. The chimeric polypeptide of claim 21, wherein the heterologous polypeptide or peptide is inserted into the CanA polypeptide at the site the subsequence was removed.

15

23. The chimeric polypeptide of claim 22, wherein the heterologous polypeptide or peptide is a 14 or a 17 residue motif inserted into the CanA polypeptide to replace the removed 14 or a 17 residue motif.

20

24. An immobilized chimeric polypeptide comprising the chimeric polypeptide of claim 1.

25. A nanotubule comprising a plurality of chimeric polypeptides as set forth in claim 1.

25

26. The nanotubule of claim 25, wherein the heterologous polypeptide or peptide comprises an enzyme.

27. The nanotubule of claim 26, wherein the nanotubule comprises a plurality of different enzymes.

30

28. The nanotubule of claim 27, wherein the plurality of enzymes comprises a biosynthetic pathway.

29. The nanotubule of claim 28, wherein the plurality of enzymes are arranged along the length of the nanotubule in the same order as they act in the biosynthetic pathway.

5 30. The nanotubule of claim 25, wherein the heterologous polypeptide or peptide comprises a chiral selection motif.

31. A nucleic acid comprising a sequence encoding the chimeric polypeptide of claim 1.

10 32. An expression cassette comprising the nucleic acid of claim 31.

33. A cell comprising the nucleic acid of claim 31.

15 34. The cell of claim 33, wherein the cell is a bacterial cell, a plant cell, a yeast cell, a fungal cell, an insect cell or a mammalian cell.

35. A transgenic non-human animal comprising the nucleic acid of claim 31.

20 36. A plant comprising the nucleic acid of claim 31.

37. A method for the chiral selection of a composition, comprising the following steps:

25 (a) providing a chimeric polypeptide as set forth in claim 12;
(b) providing a racemic mixture of the composition; and,
(c) contacting the racemic mixture with the chimeric polypeptide under conditions wherein only one enantiomer of the composition binds to the chimeric polypeptide; thereby selecting a single chiral specie of the racemic mixture.

30 38. A method for the chiral selection of a composition, comprising the following steps:

(a) providing a nanotubule as set forth in claim 29;
(b) providing a racemic mixture of the composition; and,

(c) contacting the racemic mixture with the nanotubule under conditions wherein only one enantiomer of the composition binds to the nanotubule; thereby selecting a single chiral specie of the racemic mixture.

5 39. A method for enzymatic biosynthesis of a composition, comprising the following steps:

(a) providing a nanotubule comprising a plurality of enzymes comprising a biosynthetic pathway as set forth in claim 28;

(b) providing a substrate for at least one enzyme; and,

10 (c) contacting the nanotubule with the substrate under conditions wherein the enzymes of the biosynthetic pathway catalyze the synthesis of the composition.

40. The method of claim 39, wherein the enzymes are expressed in the inner lumen of the nanotubule.

15

41. The method of claim 39, wherein the enzymes are expressed on the exterior of the nanotubule.

20

42. A cell comprising a chimeric protein of claim 1 or a nanotubules of claim 25.

43. The cell of claim 42, wherein the cell is a bacterial cell, a plant cell, a yeast cell, a fungal cell, an insect cell or a mammalian cell.

25

44. A transgenic non-human animal comprising a chimeric protein of claim 1 or a nanotubules of claim 25.

45. A plant or a seed comprising a chimeric protein of claim 1 or a nanotubules of claim 25.

30

46. A product of manufacture comprising a chimeric protein of claim 1 or a nanotubules of claim 25.

47. The product of manufacture of claim 46 comprising transistor or a circuit.

48. A medical device or an implant comprising a chimeric protein of
5 claim 1 or a nanotubules of claim 25.

49. A method for polymerizing a nanotubules comprising mixing a plurality of chimeric proteins of claim 1 in a solution comprising an iron sulfate, a manganese sulfate, a lead sulfate, a lithium sulfate, a manganese chloride or a calcium 10 chloride or an equivalent salt, under conditions wherein the chimeric protein polymerizes into a nanotubule.

50. A fluorescent chimeric polypeptide comprising at least a first domain comprising a cannulae polypeptide and a second domain comprising a 15 heterologous polypeptide or peptide, wherein the heterologous polypeptide or peptide comprises a fluorescent moiety.

51. The fluorescent chimeric polypeptide of claim 50, wherein the fluorescent moiety comprises a green fluorescent protein or equivalent.

20

52. A fluorescent nanotubules comprising a fluorescent chimeric polypeptide of claim 50.

CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND USING THEM

ABSTRACT

5 The invention provides chimeric cannulae polypeptides and nanotubules and methods for making and using them. In one aspect, the invention provides compositions and methods for the identification, separation or synthesis of proteins or ligands. In one aspect, the invention provides compositions and methods for making and using nanotubules. In one aspect, the invention provides compositions and methods for 10 the selection and purification of chiral compositions from racemic mixtures. In one aspect, the chimeric proteins and nanotubules of the invention comprise a detectable moiety, e.g., a fluorescent protein.